FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Evaluation of user interfaces for interaction with linked data in historical archives

Diogo Almeida Cunha

Mestrado Integrado em Engenharia Informática e Computação

Supervisor: Sérgio Nunes

January 30, 2020

Evaluation of user interfaces for interaction with linked data in historical archives

Diogo Almeida Cunha

Mestrado Integrado em Engenharia Informática e Computação

January 30, 2020

Abstract

In recent years, a wide variety of tasks related to ontology visualization, linked data and graph exploration have been studied and gained remarkable attention. The way data is presented, and how users can interact with it, is as important as all the complexity that is behind these systems. Therefore, it is necessary to develop research focused on the user interaction dimension. In this thesis, we aim to evaluate users' interactions with linked data systems within the context of historical archives.

The first goal of our work was to thoroughly analyze the current state of the art related to interaction with linked data. As a result, we've summarized the technologies and functionalities for interaction that can be potential ideas for our work. During the study, we noticed that in several publications related to linked data, researchers present suggestions for the navigation and interaction with data, but do not present evaluation studies with users. In the historical archives domain, we identified an opportunity regarding the study and research of interaction with linked data supported by domain ontologies.

We have worked on a project with the General Directorate for Books, Archives and Libraries (DGLAB) focused on this problem. DGLAB is currently evaluating solutions for the migrating from the hierarchical model ISAG(G) to a graph-based model. This also requires changes in terms of interaction paradigm with end-users, specifically with professional archivists. We have analyzed the current DigitArq system and developed a non-functional prototype for archivists to explore and visualize linked data with a completely new user interface (DigitArq+). New functionalities regarding the creation and manipulations of linked entities were developed, alongside with the possibility of seeing historical records within a graph perspective, a new way of interaction made possible by the graph-based format. The second part of our study consisted of the evaluation of the user's interaction with this new system.

For this study, we have developed a script for users to follow while explore the new system. We conducted a set of four sessions with DGLAB members, using the think-aloud protocol to collect information from the participants in real-time. Participants have executed three main tasks, focused on navigation, creation, and record's functionalities. As a follow-up, a final questionnaire was also made to collect more information and to better understand the advantages of the new DigitArq+ system. Finally, we have also implemented a functional web-based prototype to test a set of technologies to support the graph interaction component.

The main conclusion of our experiment was that overall, all participants indeed have an interest in a new system where it is possible to create and link data between historical records. The possibility of connecting records and entities could bring higher value to the DigitArq+ system. And also, understand what features are more relevant for each individual task, for example, the use of a graph may be useful for the navigation of records but not so friendly for record creation.

Resumo

Nos últimos anos, uma grande variedade de tarefas relacionadas com a visualização de ontologias, dados ligados e a exploração de grafos tem sido alvo de uma crescente atenção. A forma de apresentação dos dados, bem como a forma de interação por parte dos utilizadores, são tão importantes quanto toda a complexidade que está por trás destes sistemas. Torna-se por isso importante desenvolver investigação focada na dimensão da interação com os utilizadores. Neste trabalho, temos como objetivo propor e avaliar mecanismos para suportar a interação dos utilizadores com sistemas baseados em dados ligados no contexto dos arquivos históricos.

O primeiro objetivo do nosso trabalho foi analisar o estado da arte atual relacionado com interação com dados ligados. Como resultado, apresentamos um resumo das tecnologias e funcionalidades de interação com potencial para serem usadas neste contexto. Durante esta fase, percebemos que em vários trabalhos relacionados com dados ligados, são propostas soluções para a navegação e interação com os dados, mas não são apresentados estudos de avaliação com os utilizadores. No domínio dos arquivos, identificamos oportunidades de investigação relacionadas com o estudo e investigação de soluções para a interação com dados ligados tendo por base ontologias do domínio.

O trabalho aqui apresentado foi desenvolvido no contexto de um projeto com a Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB). Atualmente, a DGLAB está a avaliar soluções para a migração do modelo hierárquico ISAG(G) para o modelo baseado em grafos. Esta migração requer também mudanças ao nível do paradigma de interação com utilizadores finais, especificamente com arquivistas profissionais. Analisámos o sistema DigitArq atual e desenvolvemos um protótipo não-funcional para a exploração e visualização de dados ligados num novo sistema (DigitArq+). Foram desenvolvidas novas funcionalidades relacionadas com a criação e manipulação de entidades ligadas, juntamente com a possibilidade de visualizar registos históricos considerando o grafo associado, uma forma de interação possibilitada pelo novo formato baseado em grafos. A segunda parte do nosso estudo consistiu na avaliação da interação do utilizador com este novo sistema. Para este estudo, desenvolvemos um guião para orientar os utilizadores na exploração do novo sistema. Realizámos quatro sessões com os membros da DGLAB, usando o protocolo *think-aloud* para recolher informações dos participantes em tempo real. Os participantes executaram três tarefas principais focadas nas funcionalidades de navegação, criação e visualização de registos. Adicionalmente, foi também elaborado um questionário final para recolher mais informação sobre as vantagens do novo sistema DigitArq+. Finalmente, desenvolvemos também um protótipo funcional para testar um conjunto de tecnologias web para suportar o componente relacionado com a interação com o grafo.

A principal conclusão do nosso trabalho foi que, em geral, todos os participantes demonstram interesse num novo paradigma de interação, onde é possível criar e estabelecer ligações entre registos históricos. A possibilidade de conectar registos e entidades pode trazer maior valor ao sistema DigitArq+. Além disso, foi possível identificar que técnicas são consideradas mais relevantes para cada tarefa individual, por exemplo, o uso de um grafo pode ser útil para a navegação de registos, mas não tão interessante para a criação de registos.

Acknowledgements

I would like to express my gratefulness to all the people that supported me during these past months.

To professor Sérgio Nunes for the trust, for supervision of the project and all the time afforded.

To my family and girlfriend for all the love and comprehension! To my friends for the generous caring words during the semester! To the InfoLab colleagues, that always provided me an excellent work environment and helped me understand the fundamentals of the project in the beginning!

And finally, to all the archivists and participants that have taken their time to help me during the interviews.

Diogo Cunha

Acronyms

DGLAB	Direção-Geral do Livro, dos Arquivos e das Bibliotecas
EPISA	Entity and Property Inference for Semantic Archives (FCT Project)
CIDOC CRM	CIDOC Conceptual Reference Model
ISAD(G)	General International Standard Archival Description
RDF	Resource Description Framework
WWW	World Wide Web
API	Application Programming Interface
LOD	Link Open Data
SPARQL	Semantic Query Language
TAP	Think Aloud Protocol

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

The World Wide Web has radically changed the way we see, create and share knowledge and data across the world [6]. Along with all this information, areas such as semantic web, linked open data, and ontology systems, have been exploring more reliable ways to organize structured, in particular considering that representation and visualization can have a huge impact on how we see this information for better understandings [6]. Data visualization has become a large research field, in particular exploring the question “is there an inherent relation among the data to be visualized?”, because if data can be linked, graph-visualization and other specific techniques can play an important role in this interaction [19]. When information parts can be connected to other information parts, this can be very useful for user’s navigation, and different areas address the need for proper visualization and good reliable ways to see this kind of information.

In this dissertation, we work with the General Directorate for Books, Archives and Libraries (DGLAB), who manages the Torre do Tombo National Archive. Torre do Tombo is one of the oldest public institutions in Portugal, located in Campo Grande, Lisbon [1]. It is a national private association that manages an historical archive with hundreds of important historical records and documents. This national institution manages and stores documents to provide historians and archivists material for potential research and work. It also provides access to these records to the general public. Following a general interest in the areas of linked data and ontologies, DGLAB is working on developing models and tools to introduce the concepts of linked in historical archives.

1.2 Problem

Currently, the DigitArq system is using the ISAD(G) standard to model record information - ISAD(G) is the International Standard for Archival Description [25]. It defines a list of elements and rules for the description of archives and describes the kind of information that must and should be included in such descriptions for the specific domain. An important aspect is the fact that this model is based on a hierarchical paradigm. To introduce concepts of linked data and ontologies in the representation of archival information, DGLAB is working on a project with the goal to migrate from the ISAD(G) structure, to the CIDOC Conceptual Reference Model (CIDOC CRM) [10] model, a ontology that adopts a graph-based paradigm.

The importance of visualization definitely plays an important role when it comes to finding reliable ways to see and understand patterns in data [11]. If wrong interaction solutions are adopted, limitations can easily appear leading to misunderstandings in data, and difficulties for user interactions. With this migration we are looking to find some reliable ways to properly visualize the records represented in this new paradigm, and how users can interact with linked data. It is obvious that users are interested in friendly interfaces to have an experience as intuitive as possible, but the complexity of this demand depends vastly on the amount of data, on the ontology and how users want to see it because what is easily understandable for someone, might not be for others. Thus, we have the problem of understanding which methods to use to manipulate linked data.

1.3 Objectives

This dissertation topic argues for the need to develop a non-functional prototype for DigitArq+ system with proper visualization techniques for the navigation and manipulation of historical records, and the evaluation of the proposed solutions for the system. The need to visualize and interact with linked data can cover a huge variety of fields. Recent studies and applications, just specify techniques and tools for data visualization, but in our context there are also requirements related to the need to manipulate, edit and change data. Currently DGLAB has two applications, one for the general public to access the collections, and another for archivists to work and manage the collection. The focus of this project is on the exploration of interaction techniques for the archivists. In particular, we build and evaluate a non-functional prototype that supports access and interaction with the collection. One important aspect of this thesis is that we use proper methodologies for user experience evaluation, specifically the think aloud methodology, to extract knowledge from stakeholders during the interviews.

1.4 Document structure

This document is structured in five chapters. This first chapter introduces the theme of the dissertation, defining the context, motivation, and objectives of this work. Chapter 2 presents the state-of-the-art divided into four main parts: the introduction to the semantic web and the study of linked data strategies and technologies for web semantic; the evaluation of existing web engines for visualization and manipulation of linked data; the study of graph web engines and technologies for representation; and the study of methods for the visualization evaluation study. In Chapter 3 we present the work done during the project in four sections: the requirements for the archivists application and the evaluation of the current tool used to interact with the collection; the explanation of the new proposed prototype; the evaluation with stakeholders of the proposed DigitArq+ non-functional prototype using evaluation methods; and finally the evaluation of a post-experience

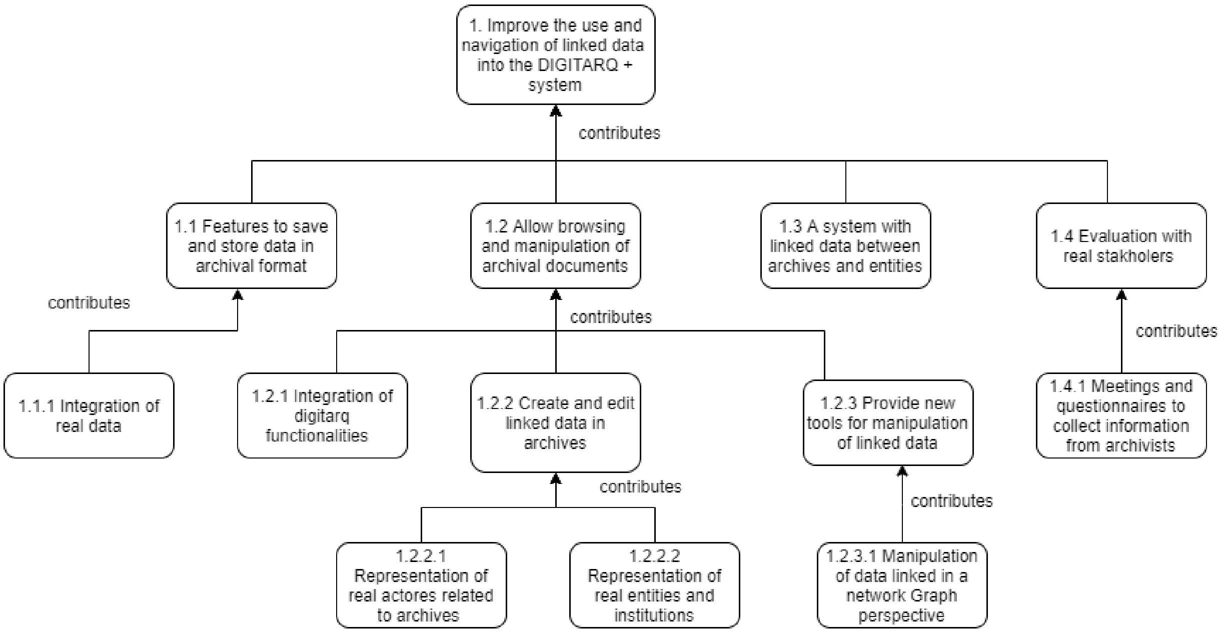


Figure 1.1: Dissertation objectives.

questionnaire for conclusions. In Chapter 4 is presented a functional prototype supporting a graph-based interaction over historical data. Finally, Chapter 5 presents the main contributions of this work, conclusions, and future work.

Chapter 2

Literature Review

2.1 Introduction

The concept of Semantic Web has gained remarkable attention in the last decade, as well as all the topics that emerged from this particular area. The representation of linked data, linked open data (LOD), and how users interact with this can provide some challenges. The availability of data and scientific publications in open access, allows these contents to be redistributed and reused, providing some relevant information [3]. So, in the following sections, we will cover the related work regarding the concept of the semantic web, linked data, visualization, and interaction techniques that have been explored. The following survey is divided into four main parts. First, we will address the state-of-the-art about LOD and detail several web applications related to this topic. Next, we will reflect on semantic web and ontology systems, and analyze two more web applications to help us comprehend some of the main features developed. In the third section, we address web tools for the development of network graphs and analyze two more web application systems that use graph visualization. In the last section, we address ways and methods for visualization evaluation.

2.2 Linked data

2.2.1 The Semantic Web

In this dissertation, we will not be developing new technologies for the Semantic Web, but it is crucial to do an overall review of the current state-of-the-art and see which technologies are here and what is essential in this area. The reason why we had to take our time to understand these technologies in the following subsections is that, even though we will develop a non-functional prototype, in future work these are some of the technologies that could help us create the prototype to interact with semantic data.

Semantic Web is known as an extension of the current World Wide Web or WWW, based on the idea of exchanging information with explicit and formal descriptions. It relies heavily on formal ontologies to structure data and perhaps to make machines understand data more clearly [5].

The challenge of the Semantic Web is to provide a language that expresses both data and rules for reasoning about the data, and that allows rules from any existing knowledge-representation system to be exported onto the Web. Technologies need to be "meaning-centered," and with the growth of the World Wide Web, the entire web community started to improve and search for technology frameworks that could help computers to make data more understandable and easier to connect.

Semantic Web technology nowadays uses formal semantics to give meaning to the disparate and raw data that surrounds us. According to the creator of the World Wide Web, Sir Tim Berners-Lee, semantic technology, together with linked data technologies, are able to build relationships between data within various formats and sources, from one string to another, helping to build context and creating links out of these relationships. These technologies include tools for auto-recognition of information, topics and meaning-extraction, categorization, and so on.

Therefore, we can presume that one of the main goals of the Semantic Web is not only to train machines to behave like people, but to develop technologies and languages that make information readable to machines. The aim is to develop a technological model that enables machine-assisted global knowledge to be shared across computers.

Important technologies to reference in this context are: the Resource Description Framework (RDF), integration of eXtensible Markup Language (XML), web ontology language (OWL), computational agents, semantic query language (SPARQL), Really Simple Syndication (RSS), N-Triples, among others, will all contribute to the emergence of a web of data that contributes to the interoperability and cooperation between computer systems.

2.2.1.1 Semantic Web technology concepts

In order to expose data that is generally hidden on the Web, Tim Berners-Lee [22] references five technologies that helped the Semantic Web to move further and that are still being used in today's society.

The first technologies mentioned are the Identifiers, URIs, and URLs. The most common these days are the URLs, which are used to identify the address of a web page. An URL basically consists of an expression that can identify a computer and a specific domain, where the web document resides, the file name of the page that is being visited, and the virtual file directory where the webpage is allocated. Uniform Resource Identifier (URIs) work in the same way but in this sense are used not as the address of a point, but as resource identifiers. So, this way, an URI can be defined as a string of characters that identifies a particular resource in an unambiguous way. To make sure that all resource identifiers guarantee uniformity, all URIs follow a predefined set of syntax rules. Such identification enables the interaction with representations of the resource over the network, typically in the World Wide Web, using specific protocols. Although the syntax to create a URIs is carefully governed by the Internet Engineering Task Force (IETF), the actual control over URIs is decentralized, so that no one person or organization controls who makes them or how they are used, meaning therefore that no permission is needed to create an URI.

The second technology mentioned are markup languages, which help the web community define the internet structure, such as XML, HTML, SGML. Nowadays, one of the most used technologies is HTML [7]. Semantic HTML is the use of HTML markup to reinforce the semantics or meaning of the information in web pages and web applications rather than merely to define its presentation or layout style. The Semantic HTML is traditionally processed by the web browsers, and also as many other user agents. Semantic HTML defines specific tags to indicate clearly what role is played by the content that those tags contain. That explicit information helps computer systems to understand which content is essential, which is a subsidiary, which is for navigation, and so on. By adding semantic HTML tags to Web pages, it provides additional information that helps computers understand the roles and relative importance of the different parts of any user's page.

Another technology mentioned by Tim Berners-Lee is XML [4]. XML was designed as a simple way to store or send documents across the Web, which allows developers to add meaning to the data being stored or transmitted. This functionality is made available by allowing a developer to create meaningful tags that contain any data. When the XML file is then interpreted, a computer application can parse the tags and perform certain functions on that data as determined by the content and attributes of the tag, which encloses it. XML also allows for namespace declarations within each tag to hold URI information, thereby ensuring that name tags created by one person do not conflict with those created by another person and making it the perfect mechanism for the Semantic Web.

Another technology also mentioned by Berners-Lee is the Resource Description Framework (RDF). There has ever been some misunderstanding between the relationship between RDF and XML. The main difference is that XML is syntax, while RDF is a data model. The RDF Standard or Resource Description Framework is a W3C recommendation and one of the core standards of the Semantic Web. The RDF encloses a standard of ontologies, for the description of any type of Internet resources such as a website and the content itself. RDF actually sets standard metadata to be embedded in XML coding, and its implementation is exemplified by the RDF Schema, or RDFS, which is part of the standard specification. The idea of RDF is the description of data and metadata through a "triple" value recursion-property scheme, and a coherent way to access Web-published metadata standards. In other words, it represents a data model that can be compared to a relational model. RDF is just another way of reorganizing data but as a graph perspective. RDF can have two things that are related in some way through a link that can connect the entities and data.

Although there are some other existing technologies in the Web community, the fifth technology mentioned by Tim Berners-Lee is the Ontology. An ontology basically refers to a formal definition of the relations between terms within a specific domain. The typical Web ontology consists of both taxonomy and a set of inference rules. The taxonomy defines all the classes of objects and any relationships between them. The inference rules allow an application to make decisions based on the classes supplied without needing to understand any of the information provided.

The use of these ontologies may be suitable for computers because two different databases may use completely different identifiers to identify the same aspect or concept, for example, the last name and the surname. A Web application that wants to compare these two concepts needs to know that these two terms are being used to mean the same thing. In order to do this, computers need to have a method of discovering such common meanings for whatever databases it queries and not depending on the language.

2.2.1.2 Conclusion

Overall, the Web community has defined several ways to represent different scenarios or situations and make them possible to understand by machines. Ontologies and RDF schemas are used to organize data in a structured way so that computers can understand and process the data being published. The potential of the Semantic Web to solve real-world problems in inter-device communication and to be able to classify information is tremendous. Unfortunately, to achieve these results, it is necessary to understand that its power is more applicable to certain types of information than it is to others [11]. In the next following section, we will address another aspect related to linked data and which methods are adequate for visualization in this context.

2.2.2 Linked data search systems

2.2.2.1 Introduction

In the previous section, we have mentioned how Semantic Web can cover a wide variety of sources, technologies and make data available online, but it isn't just about that, it is also about providing interfaces so that computers can know and explore data on the Web. Linked data or linked open data (LOD) is about using the Web to create links between data from different types of sources. While the Web uses primary units like HTML to connect documents through hyperlinks, linked open data relies on the use of documents containing data in the RDF model.

In this following section, in order to comprehend how linked data is really important to help computers to understand how data can be linked and visually understandable for users, we will list some current work related to search systems for linked data research. Once again, we will discuss another publication by Tim Berners-Lee [22], where he mentions linked data topics about the Web, applications, projects, and how this data can be applied in many different forms.

2.2.2.2 Marbles linked data web application

The first web application mention by Tim Berners-Lee [6] was the Marbles system and how it shows the information related to a single entity. Marbles, based on RDF web technology, is a server-side application that formats Semantic Web content for XHTML clients using Fresnel lenses and formats [27].

For users to understand the origins of the information, Marbles displays colored dots in front of the data to show the correlation between the data shown and to his sources. Relevant aspects of the Marbel's example in Figure 2.1: the data that users see is collected from multiple sources and integrated into a single graph (not visible) that persists during the users' session. When Marbles is provided with the URI of a resource to be displayed on the interface, it attempts to refer to it. In

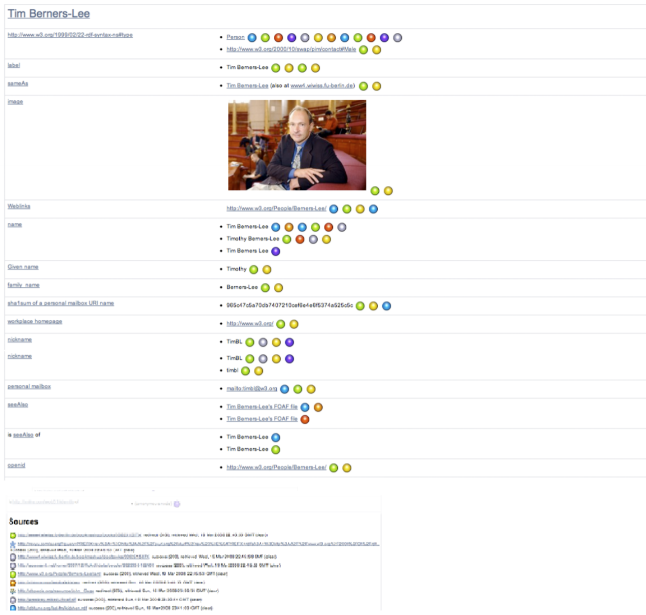


Figure 2.1: Tim Berners-Lee's Marbles profile.

parallel, it queries 'Sindice' for data sources that contain information about the provided feature and 'Revyu' for some revisions of the information. About the visualization of the linked data, apart from a full-details view that lists all known properties for a resource, Marbles allows for the generation of Fresnel-based views that are complemented by corresponding CSS stylesheets and presenting the result in a table, with multiple rows. Overall, even when using linked data over an RDF, we can see that not always a graph-oriented view is the only solution to give the idea of other links connected, and adding colored dots to correlate information was a solution to approach and resolve the problem.

2.2.2.3 FALCONS linked data web application

In the traditional hypertext web, browsing and searching are often seen as the two dominant modes of interaction (Olston and Chi, 2003). While browsers provide the right mechanisms for navigating the information space, search engines are often the place at which that navigation process enrolls. Several search engines have been developed that crawl linked data from the Web by using and follow RDF links. Regarding these services, they can be divided into two categories: human-oriented search engines and application-oriented indexes such as Falcons [6].

Search engines like Falcons provide functionalities like keyword search services oriented towards human interactions, similar to Yahoo and Google. Users are presented with a search box in which they can enter keywords related to the item or topic they search, and the application will return a list of results that can be relevant to their queries. However, instead of simply providing links from search results to the source documents in which the entered keywords are mentioned, Falcons provides a more detailed user interface that explores the underlying data structure. A

Figure 2.2: Falcons web application.

traditional list is shown, along with additional structured web crawled data and links to related entities.

Falcons supplies users' three options in the top bar, Figure 2.2 option to search objects, concepts, and documents, each of which leads to a slightly different presentation of the results. While the object search is suited to search for people, places, and other more concrete items, the concept search is oriented to locating classes and properties in ontologies published on the Web. The document search feature provides a more traditional search engine experience, where results point to RDF documents that contain the specified search terms.

Overall, the representation of the linked data in Falcons doesn't seem to be well organized, and once again, it relies only upon rows, with multiple colors for each hypertext link [22], and each row has a link to redirect to another hypertext web page.

2.2.2.4 Wikidata: free collaborative knowledge base

In continuation of linked data systems, the two previous ones were proposed between 2000 and 2004, but many more systems have been recently developed related to semantic web and linked data web applications, which we are still using today, such as Wikidata.

The system Wikidata is a free web-based open knowledge-base that both machines and humans can read and edit. Wikidata acts as a storage of information for the structured linked data of other web pages from the world of Wikimedia, such as Wikipedia, Wiktionary, Wikisource, and others. All content and data of Wikidata is available as a free license, which makes it easier for computers and users to interlink with other open datasets in the linked data web.

Since Wikidata is a recently linked data web application, we will take a closer look at some of the points and evaluation techniques presented on his system.

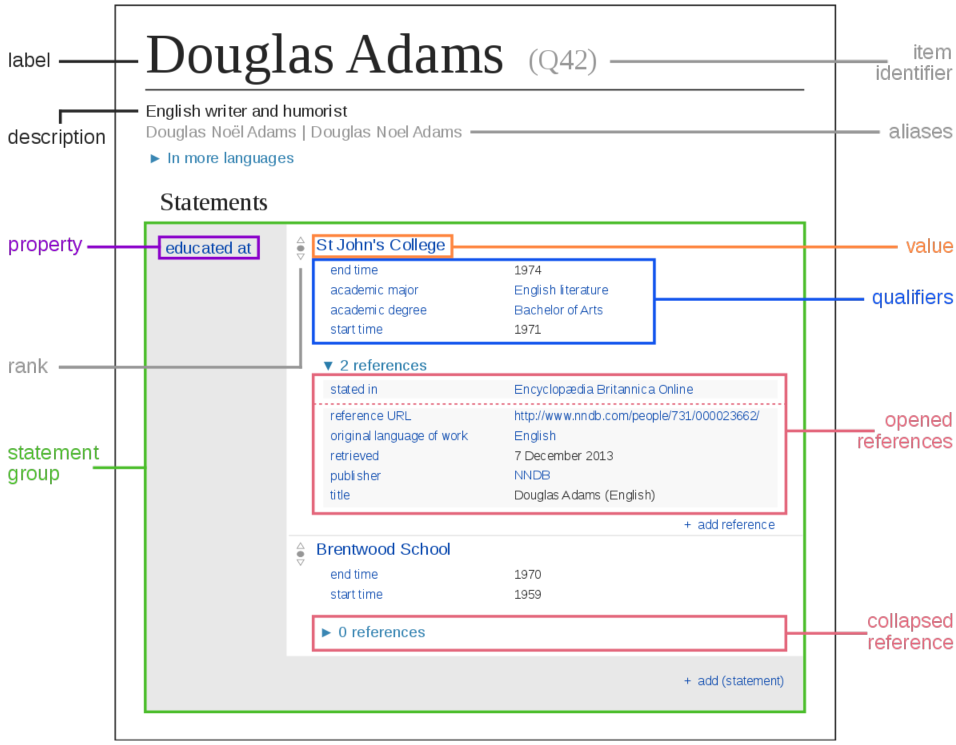


Figure 2.3: Wikidata page: Douglas Adams.

In the following subsection, we will address some figures for better understanding and mention some important aspects of evaluation of interactions with link data.

Wikidata, like Wikipedia, is organized in pages, and this is also how the data is structured behind the system. Every subject on which Wikidata has structured data is called an entity and every entity has its page [14]. The system can be distinguished by two types of entities: items and properties. In more familiar terms of semantic technologies, items can represent classes and individuals, and the Wikidata properties resemble on the resources description (RDF) properties.

Items are used to represent all the things in human knowledge, including topics, concepts, and objects. For example, the "1988 Summer Olympics", "love", "Elvis Presley", and "gorilla" are all items in Wikidata. A property describes the data value of a statement and can be thought of as a category of data, for example, "color" for the data value "blue". Properties, when paired with values, form a statement in Wikidata. Properties are also used in qualifiers. In Figure 2.3 of Douglas Adams, is represented and distinguished all these properties.

Virtually every Wikipedia article in any language has an associated item that represents the subject of his article. Every item has a page where users can view and enter the data. For example like in Figure 2.3, the item page for the English writer Douglas Adams [1] , each title alongside with the name each page has a code, similar to "Q42" rather than "Douglas Adams" since Wikidata is a multi-lingual site. Therefore, items are not identified by a label in a specific language, but by an opaque item identifier, which is assigned automatically when creating the item and which

[1] Wikidata: Douglas Adams

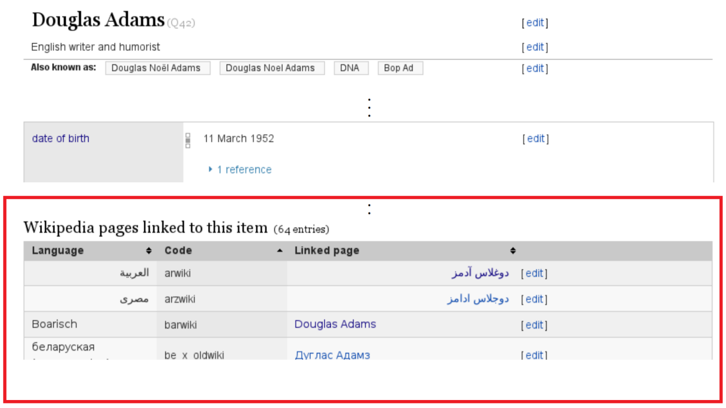


Figure 2.4: Statement of Douglas Adams with multiple languages in linked data.

cannot change later on. Item identifiers always start with "Q" followed by a number, allowing the Wikidata to connect data across all this platform with unique identifiers.

Every entity's page can be divided into three parts of identification: labels (name), descriptions, and statements, where statements are where all the linked information normally is. As in Figure 2.3, these parts are: the first label with the name "Douglas Adams" following with the unique code (Q42); below, a short description "English writer and humorist"; under it, a list of aliases and some other names that are related to the same person; below, a list of statements (the richest part of the data); the list of site links, are links to pages about the item on Wikipedia and other projects that are related and mention the item.

About the language of the items, each item can have his own terms in every language that is supported by Wikidata, but the primary language that is displayed on the browser depends on the user's settings.

An example in Figure 2.4 site links can be given for any of the 286 language editions of Wikipedia, and several sister projects, such as Wikivoyage and Wikimedia Commons. Site links are functional and inverse functional (one item per site link, at least). In opposition to the former system of Wikipedia language links, site links should only be used for articles that are precisely about the item and not about any other related topic. Wikidata prevents from having different pages for the same item, by displaying site links in many different languages at the same column. On other occasions, some items do not have any site links, for example, the item "female" with code (Q6581072), which is used as a possible value for the sex of persons. As shown in Figure 2.5, the result doesn't show the representation with all the data linked to this particular item. Instead, it shows potential and similar results related to this keyword.

The result for 'Douglas' in Wikidata has a huge list of 'statements'. Some may be as simple as a column with one line of text, and other statements can use some specific qualifiers making the interaction and the interface more complicated. As Fredo Erxleben mentions [14], in this example we address Douglas 'spouse' according to his publication that refers to qualifiers and links.

The entire data model presented in Wikidata's statements is more complex than Figure 2.4

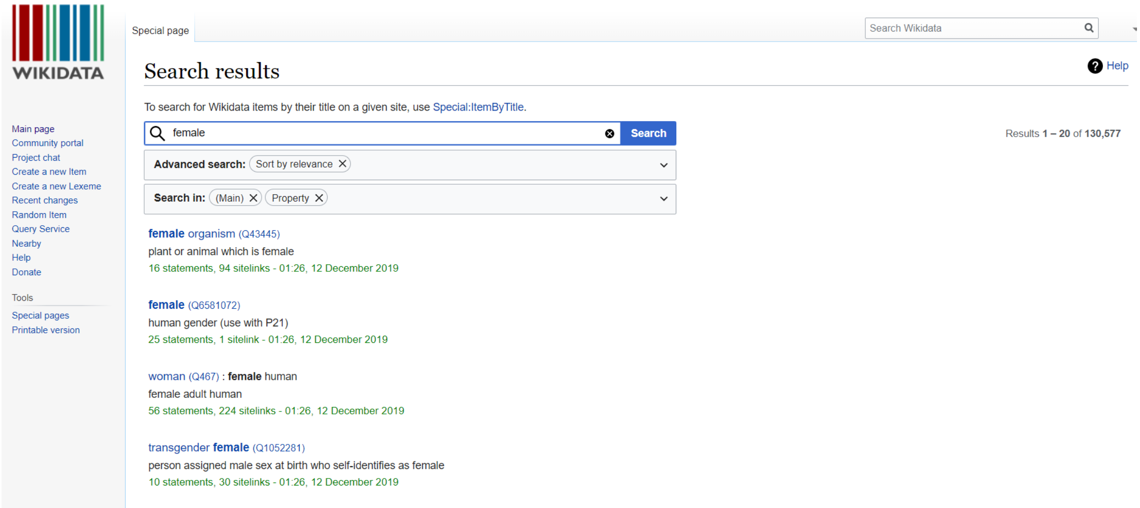


Figure 2.5: Wikidata page: Female Result List.

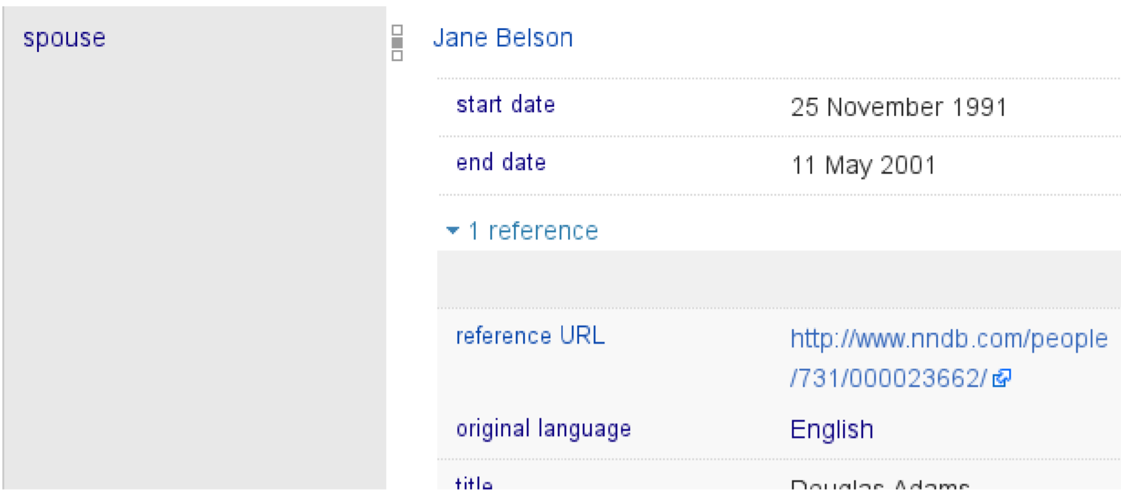


Figure 2.6: "Spouse" statement of Douglas Adams.

suggests. On one hand, the wiki statements can be represented with qualifiers, which provide additional context information for the entity we are looking for. On the other hand, every statement can include one or more references, which supports the statement. A statement where both aspects are given is shown in Figure 2.6. This following example shows exactly that, a statement that has more context information inside it, besides only links and text to other hypertext links, it uses Wikidata qualifiers. In Figure 2.6, the main value pair in this statement is the Douglas' spouse, "spouse: Jane Belson" (value P26: Q14623681), which has additional context information inside the content box. Each statement can have 'qualifiers'. Qualifiers are generally used to describe the real value of a property given inside of a statement. Qualifiers are used in various ways in Wikidata, for example, specifying the validity time of a claim is the most common usage today. An example of a claim, in Figure 2.6 the "start date: 25 November 1991" and the "end date: 11 May 2001," indicates that Adams and Belson have been married since 1991 and until his death in 2001. And to do such representation, we use properties like 'start date' (P580) and 'end date' (P582) of type 'time'. These particularly pairs of property-values refer to the central part of the statement and not to the item itself about Adam Douglas.

Besides qualifiers, Wikidata uses many other kinds of annotations that provide contextual information for items and statements. This kind of annotation and evaluation of existing ideas, later on, the dissertation will help us to understand better potential ways for linked data interpretation and data connection.

2.2.2.5 Conclusion: overall linked data visualization interaction

The vision of the Semantic Web has been interpreted in many different ways [23]. However, despite this diversity in interpretation, the original goal of building a global Web of machine-readable data remains constant across the original literature on the subject. According to Sir Tim Berners-Lee, “The first step is putting data on the Web in a form that machines can naturally understand, or converting it to that form. This creates what I call a Semantic Web" [22]. The development of a Web of data that can be processed directly or indirectly by computers. Therefore, Semantic Web can be established as the overall goal or result of a process, while linked data provides computers and researchers the means to reach that goal.

2.2.3 How semantic web reflects on ontologies

As we move further in this dissertation, we notice that several topics need to be addressed. Since we'll work on an archival system that works alongside with ontologies, it will be essential to do some research and analyze the science behind it. Ontologies are a popular topic for many communities of researchers. In Computer Science, the representation of an ontology embraces a formal way of naming and defining properties and relations between concepts, data, and entities of the domain. Meanwhile, in Philosophy, with a similar definition, an ontology can also be the study of existing things and what kind of entities are represented in the universe. It represents the study of metaphysics' information, gathering knowledge for better understanding the essence of things [20].

Although ontologies can be defined in different ways for different areas, both Computer Science and Philosophy share the idea that ontologies represent entities, events, and ideas, with all their properties and relations, according to their categories of a domain. The connection between these properties and entities that represent a specific domain aims to gather information and to represent knowledge. But in both fields, there are still some considerable work and problems of ontology engineering to exactly understand what the representation of these domains and an ontology is possible to do.

In the field of Philosophy, authors say that "ontologies carve the world at its joints" [20]. This means that ontologies provide a domain conceptualization and not a structure for a data container. One good example mentioned by Chandrasekaran [8] is that, for the same ontology system, when changing from one language to another, we only need to change the language itself and not the structure, since the conceptualization stays the same. In the meantime, ontologies have grown beyond philosophy, and now we can easily see them being used in any area of our society, even in this technological world. No matter the area or purpose, ontologies can help us to create domains in multiple aspects, but in this project, we address their use in an archivist environment.

Nowadays, historians and archivists can study digitized documents that are available online. These collections of materials available can adopt ontologies, helping archivists to organize documents in a well-defined domain [2]. For the past decades, archivists and researchers have conducted research on visualization solutions, since these techniques have an appealing potential when it comes to create, explore or verify complex and large collections of data such as ontologies.

After addressing in the previous section some web applications related to linked data, in this next section, we will describe some web applications relatively to ontologies and linked data in an archival environment that is exactly our goal for the future work on this thesis.

2.2.4 Exploiting web applications for ontology systems

2.2.4.1 Museo del Prado: archival web application

The Museo del Prado or officially known as "Museo Nacional del Prado" [18] is the main Spanish national art museum, located in the center of Madrid. It is considered one of the most prestigious museums for having the finest collections and history of European art, from the 12th century to the early 20th century. The rare and important collections presented in the museum attract people and historians from all over the world. Nowadays, the Museo del Prado showcases more than 8,200 drawings, 7,600 paintings, 4,800 records, 1,000 sculptures, and many other related works and historical documents.

Next, we will address some functionalities that may be important for future work such as, how documents are connected, how images and records are related to the author, and a few more aspects that will help us further in this dissertation.

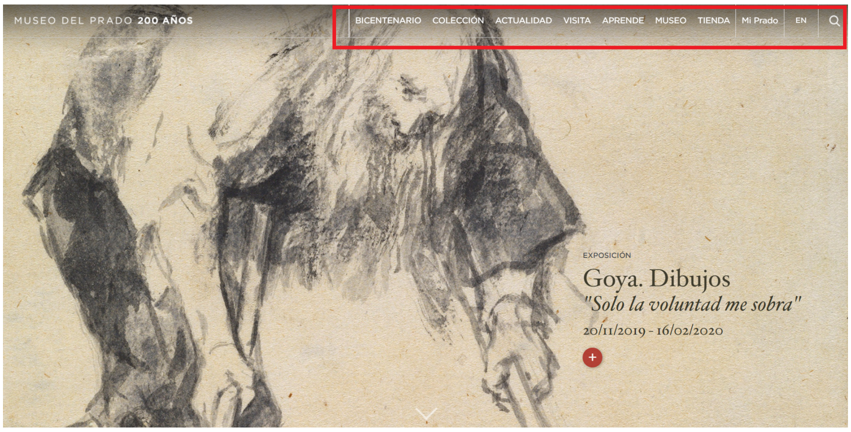


Figure 2.7: Museo del Prado Web Site.

In Figure 2.7 is the first impression users have in terms of interaction, with a list in the navbar with all the features: bicentenary/bibliography, collection of documents, what's on, visit, learn, museum and shop. These are the main functionalities presented.

In this thesis we are focused on ways of representing data and linked objects such as records, authors, institutions and any other data related. In Figure 2.8 we have the two possible ways users can see information after clicking 'collection'. The techniques that Museo del Prado uses to present data are really interesting and user friendly, they include two distinct visualizations. Figure 2.8 on the left side we have a list organized in line with square images and information related to it (A), then (B) and (C) are different ways to filter the same data to have a more accurate result. On the right-hand side, this navigation technique was mentioned by archivists in one of the meetings with DGLAB. Museo del Prado provides users with a very interesting way of navigating in the documentation, specifically a timeline navigation. This approach may be interesting when it comes to the interaction of linked data since we have information about what a document is connected to.

The way documents are presented and data shown are of great relevance for future sections on our dissertation. In Figure 2.9 is possible to see on the left side section 'A', Museo del Prado has a section to view the image and users can zoom in, zoom out and move the entire document. In section 'B', related documents are listed below the image. An important piece of information to detail for future work is how they represent their documents and what kind of information they use. In the related section 'B' The Museo del Prado uses an image, title and year of the document were created. How this linked data is shown plays an important role as we'll see further on this dissertation for linked data related to the records. On the right side 'C' is where is listed all the text and information related to the records or paintings.

For the second way of visualization during navigation, in Figure 2.10 in the timeline, after double-clicking an element, a modal with a short description of the document is displayed. This technique has its advantages and disadvantages, but it may be a technique to consider for the representation of entities and actors.

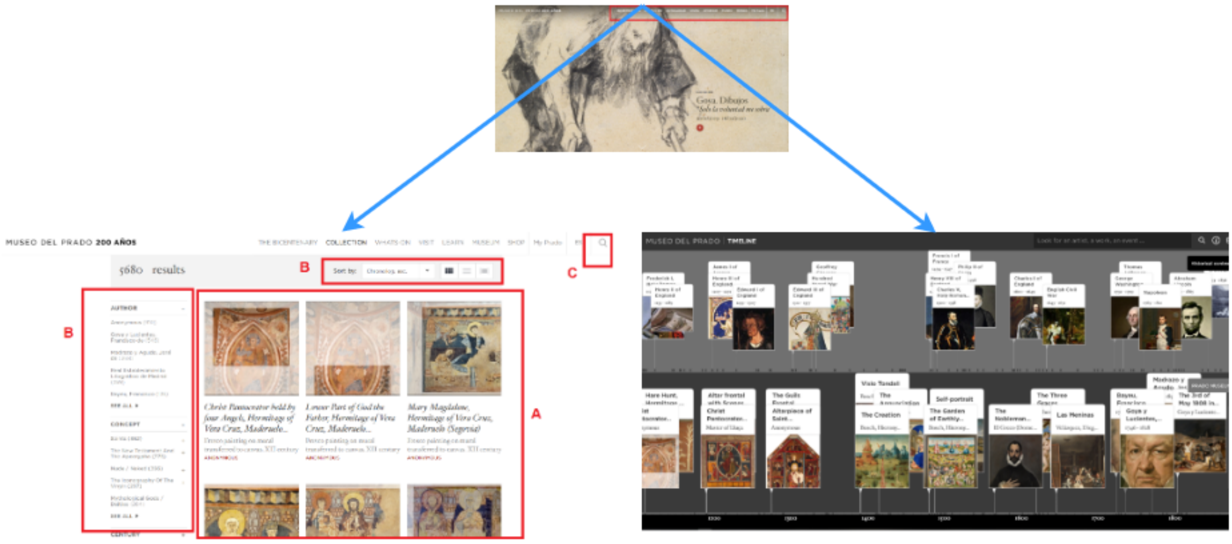


Figure 2.8: Museo del Prado: Documentation presentation.

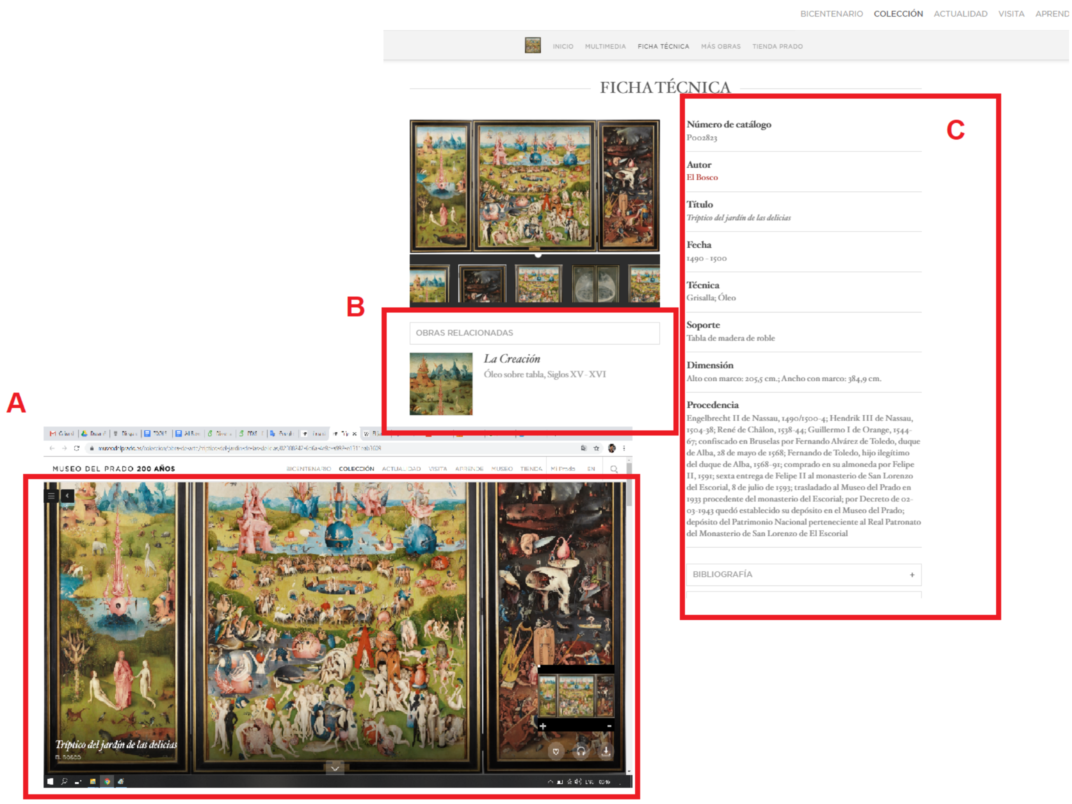


Figure 2.9: Museo del Prado: Document Information.

Figure 2.10: Museo del Prado: Modal with Information.

In general, important aspects to take into consideration for future work are these types of navigation with suggestions and ideas on how to handle historical data and also how other institutions represent national records, in particular in the way they are linked to other entities.

2.2.5 DigitArq web application analysis

As mentioned before, DigitArq is a software system managed by DGLAB. DGLAB aims to meet the needs of a diverse audience seeking to access a multitude of information and services. The DigitArq system provides two applications, a desktop application targeted at professional users (i.e. archivists), and a web application for the general public. The DigitArq Desktop [16] will be addressed later on this dissertation. The DigitArq web will be discussed in the next subsections.

2.2.5.1 Site map view

Presented below we have a sitemap that describes how the web application currently works based on the web pages available for non-authenticated users:

As we can see in Figure 2.11 the website currently has 6 ramifications and about 20 pages online. At the home page, users can easily get access to what pages are available and, most importantly, they can write any specific data to search on the platform. The simple search path provides users the ability to check for documents and records, the most viewed, and recent ones. Advanced search works similar to the previous one, but with filtering tools providing a much more detailed search (will address this later on this section). Highlights are offered to emphasize prioritized documents. A help page is also available for regular users and finally, online services and login are features available for users that can be logged into the system.

In the following subsections, we'll address all these interface features, discuss how these pages are connected and see how users can interact with the system.

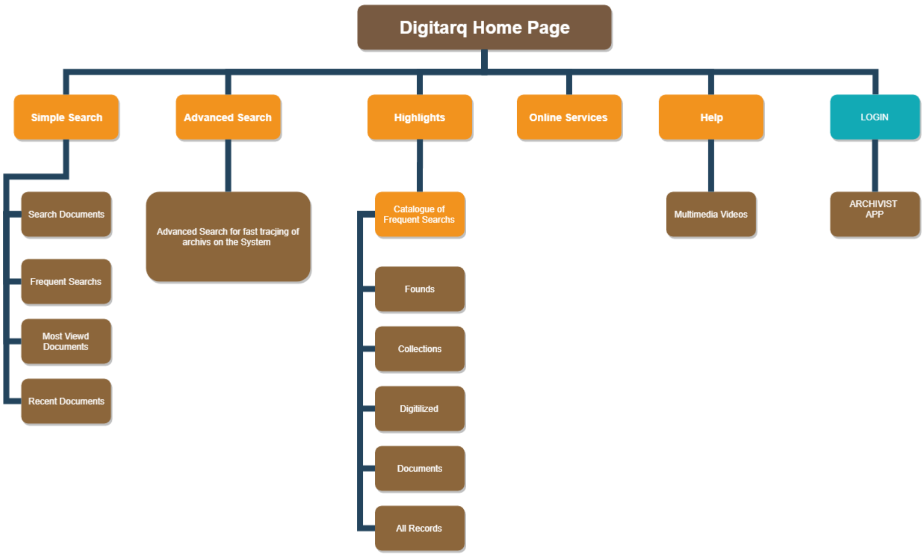


Figure 2.11: DigitArq web application.

2.2.5.2 Interfaces and common features

Before we go deep into some major functionalities of the system, we've decided to explore some principal and common features of the web app and see how it is organized.

In Figure 2.12 are presented both interfaces for mobile and Web interfaces, and as we can see in the Main Page it is organized into four sections on both devices: DigitArq logo, page header, main content, and footer.

The first one represents the national archives from Torre do Tombo (DGLAB). The second layer, as we can see in Figure 2.11 is where all the other possible actions can be accessed by users. The "Content" section is where users can directly search for records.

In this research, we can notice that the mobile application is not so reliable and has some possible changes that need to be made. The fact that the entire page is not completely full may lead to some bad experiences. This aspect will be taken into consideration for future work. In the following subsection we will address possible user experiences on the DigitArq web application by doing storyboards to identify navigation patterns.

2.2.5.3 Current web DigitArq wireflows

For each subsection, we will first explain the navigation and how users can interact with the system in an organized way. After seeing Figure 2.12, specifically how the interface looks like, and where the main content is located, we've decided develop a wireflow view of the system. Wireflows are divided in two main actions: navigation through the website (in Figure 2.13) and search for documents (in Figure 2.14).

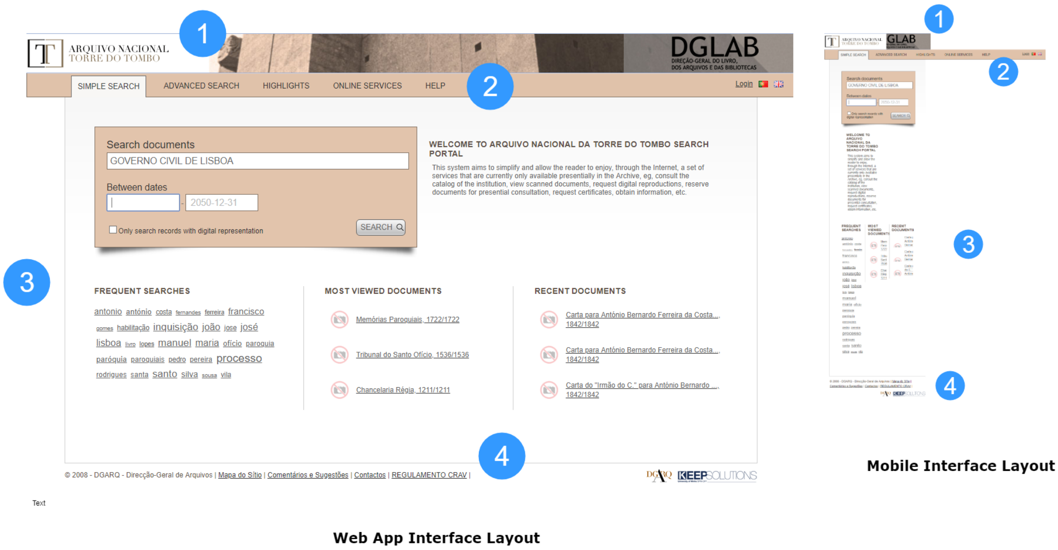


Figure 2.12: DigitArq Web application.

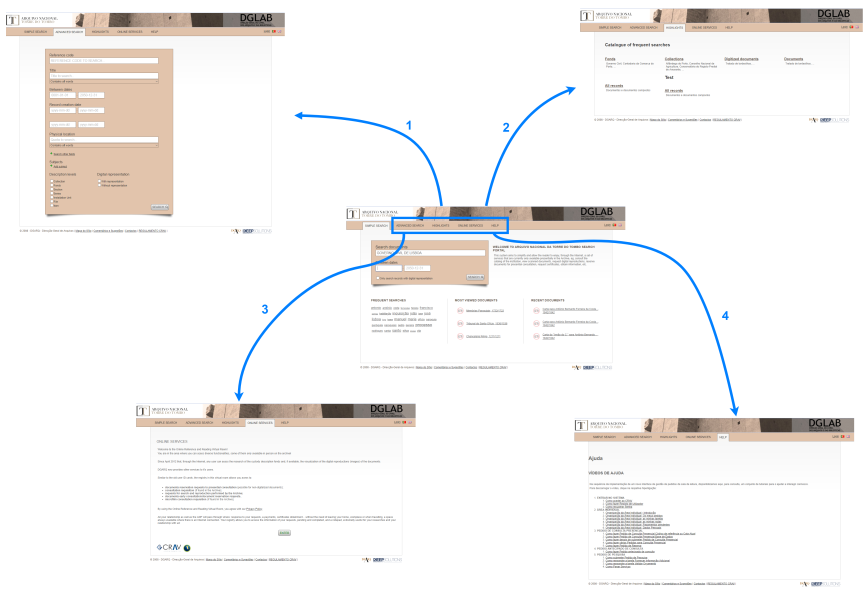


Figure 2.13: Wireflow centered on the user's options.

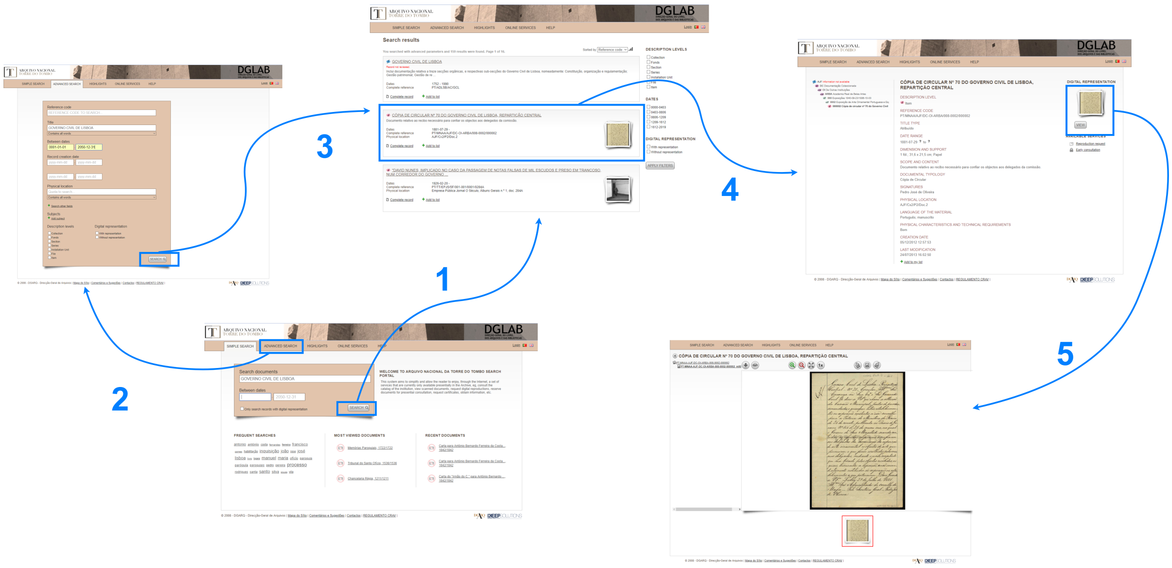


Figure 2.14: Wireflow centered on search for records.

2.2.5.4 User's actions in the first page of the system

Regarding the user's intention, each page has its own functionality and to better understand the wireflow, in Figure 2.13 we have all the actions that users can take from the main page to the other individual pages on the website.

In **Action no. 1** users will get into the "Advanced Search" page. In this page, users are provided with tools to make their search more specific by introducing filtering parameters (all listed below).

With **Action no. 2** users get into the "Highlights" page, here DGLAG provides users with the most frequent searches on the website.

Action no. 3 is only available for registered users.

With **Action no.4**, users have access to a help page with links to videos explaining how to use DigitArq web.

2.2.5.5 User action for search and navigation of records

Observing Figure 2.14, these are the following steps users should take in order to execute the search, see the record's information, and the associated image.

Action no. 1: When users are on the home page, they can quickly search for records by providing the name of the document and/or the date range (in numbers), and the website will show users all records that matched these parameters.

Action no. 2: If users want to do a more specific search using other parameters to filter data, action no. 2 lets users access "Advanced search" and add specific filters to the search execution. The evaluation of this list of filters, presented in Table 2.1, will help us decide which parameters we should take into consideration, alongside with the requirements from the users, to use and implement in the prototype.

Filter	Functionality
Reference code	Is the key code of the archive. Goes directly to the document.
Title	The title of the document
Data Ranges	Provides all documents in a certain period of time
Physical location	Text where the archiv was created
Search other fields	List of more specific paramenters to search for archivs
Add subject	Users can filter the records by predefined cities
Digital representation	Visualization only of records with a digital image
Select description level	Filter to select the type of the records

Table 2.1: Filters available in the DigitArq web

Action no. 3: After collecting these criteria, with action no. 3 users will have access to a list full of documents and records that are compatible with all the previous parameters. On this page, users can still filter the archives with description levels, date ranges, digital representation, and sort the archives by date, title or reference code. (All definition of each tool, defined in action no. 2)

The main list provides short information about the records (blue rectangle in Figure 2.14) such as the title of the records, the records date, the reference code, and its physical location.

Action no. 4 After the user selects the document he wants, the action no. 4 will take users to a page with all the information related to it. On the left side, users can see how the information is organized hierarchically, with all the fields and ISAD(G) tags. The middle section is where we have the main content with information related to the archive, such as the reference code, description level, title, type of document, date range, etc. On the right side of the page, if images are available, users can click to get access to the digital representation of the records (blue square on Figure 2.14).

Action no. 5: After getting access to the digital representation of the records, users can see a real image of the records, with features to zoom in, zoom out, rotate and download the image.

2.2.5.6 Conclusions and evaluation

The DigitArq web system is designed for users who aim to search and access historical records. The evaluation of some of these features will somehow help us determine potential tools and functionalities that should be addressed during the creation of the prototype. Even-though we will work with the DigitArq desktop application, having an overall perspective on how both applications interact is as important as the evaluation. Some of the techniques used and filters provided will also be a part of our prototype.

	Visual Index	Interactive	Languages
Arbojs	1	Yes	Js
Sigmajs	3	No	Js
D3	+100	No	Js
Alchemajs	8	Yes	Js
Visjs	60	Yes	Js
Plotly	15	No	Js,R,Py
Graphviz	37	No	Java,Ruby,Py,Js, Perl, php, tcl, guile
ReactNative	+20	Yes	JS

Table 2.2: Web graph tools for development.

2.3 Graph data visualization

2.3.1 Introduction

Nowadays, in the world of semantic web and linked data, visualization tools and technologies are essential to analyze massive amounts of information and make data easily available for decisions. System tools and mechanisms have been developed during the past few years to provide better interfaces between computer systems and users. The amount of data related to any specific domain can be large enough, leading to a bigger concert about their visualization. Graphs are a traditional and powerful tool for representing data sets and their relations.

Being able to visualize this data as graphs can provide even to non-experts, an excellent way to understand the data, the content itself, identifying potential patterns in data, and many other characteristics [21]. So, why is data visualization so important? In a short explanation, users prefer quick and simple interaction techniques rather than read text or a spreadsheet, making data easier to understand.

The visualization and representation of the collection in a graph-based engine would be a possible way to interact with the data. In summary, the following section will address two topics: the study of graph-based solutions that may help us in the future work for the development, and the state-of-the-art related to web systems that provide the visualization of linked data using graph-based strategies.

2.3.2 Libraries and frameworks supporting graph-based layouts

Since one of the goals of this thesis is to develop a web-graph prototype capable of managing, visualizing and exploring historical data in a graph, we have listed in Table 2.2 some tools that can help us to visualize graph-based information. Important aspects to take into consideration when choosing a graph technology would be the number of visualizations techniques (Visual Index) available for each technology; if the technology provides interactive features for the graph (Interactive); and another important aspect for developers the programming language that all technologies support (Languages).

Arbosjs is a graph visualization framework built with web workers and jQuery. It provides an efficient force-directed-layout algorithm and it provides developers for screen-drawing by using canvas, SVG and HTML elements. Sigmajs is another drawing software tool, it provides a lot of built features by using Canvas e WebGL to render the mouse movements. The default configuration of Sigmajs deals with mouse and touch support, refreshing and rescaling the container's size. Making it possible for developers to do their own functions. D3 is another JS library that allows users to visualize graphs on the web. It makes use of SVG, HTML5 and CSS technologies. D3 library also provides more graph layouts than the others, which can be good, as it gives developers more leeway. Alchemyjs is a dynamic browser-based visualization library. It was defined to be easy to use and together with Neo4j. This tool is also known for its well-designed layouts and user interfaces. Plotly is another interesting tool since it was built using library D3.js. This way, it not only provides visualization for JS but for several programming languages (as we mention in Table **3.4**; Graphviz is an open-source graph visualization software. It has important applications such as in networking, bioinformatics, software engineering, database and web design, machine learning, and in visual interfaces for other technical domains. ReactNative can also be a relevant technology because one of the latest software releases was the ReactNative-Vis, where programmers have available charts and interactive graph layouts to use. Alongside with great documentation, ReactNative has numerous examples in codepen which can help developers to get involved with the software.

In the following subsection we will present some web-based systems and projects that use graph-based solutions for navigation in linked data. The point of this evaluation is to see what kind of techniques and, especially what functionalities, are available for the graph visualisation and manipulation techniques.

2.3.3 Overview of solutions using graph-based layouts

2.3.3.1 Heritamus

The first web application we would like to mention is Heritamus. Heritamus is a web-based application for organizing, structuring, and retrieving historical and ethnographic data on heritage, contributing to overcome the asymmetrical representation of knowledge by bringing practitioners into the core of the research process [15]. Heritamus assumes that the community and its knowledge can be traced by a network of dynamically connected "nodes". Users simply have to identify the items that Heritamus recognizes as their heritage (tangible and intangible) and wishes to input in the graph.

Heritamus has 2 options, one for regular users, to see data in a graph layout, and other for authenticated users, to create a graph.

In Figure 2.15 is displayed the first impression and interaction with the system both for regular and login users. The system represents nodes and links. Each node represents an entity, which can be an album, concepts, music, song, document, group, melody, object, actor, poem, painting, phonograph, and customization. In the upper left side are listed, with colors, the meaning of

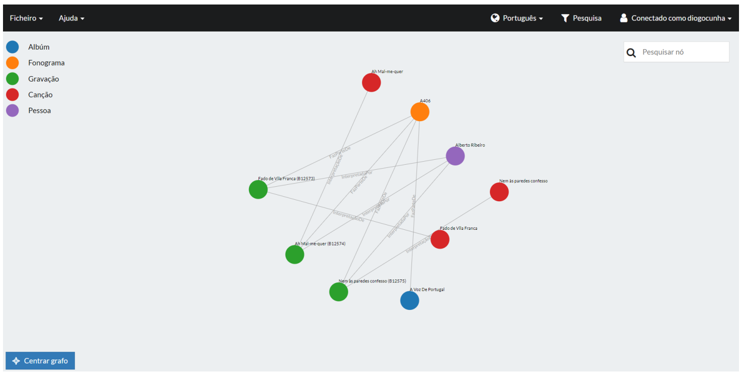


Figure 2.15: Heritamus network graph.

each node, 'A' in Figure 2.15. For the links, they are represented as the type of connection. If the document 'belongTo' someone, or if the music is 'performedBy' someone. For each link, Heritamus has a relation type for the connection of nodes.

If users want to know more about a certain node, a modal is opened with double-click (Figure 2.16). On the modal, it shows more information about the relations on the node.

2.3.3.2 Login users for graph creation

After logging into the system, users can create their own graphs. In section 'A' in Figure 2.17 we can write the name of the node we want. If the node exists, the system will list nodes that have matched the name, if it doesn't exist Heritamus will create a new node and display it as shown in section 'B'. After in section 'C' we added 2 more new nodes and finally, we can add information to our new blue node.

For the links, this was a challenging task using the Heritamus platform. Although the creation of nodes is quite susceptible and even easy to interact with, when creating a new graph, the link between entities was not as intuitive as that. After some searching and browsing, it was not possible to find any way on how to create and connect nodes. Even after a closer look at the help search page, no documentation or script was found. The fact they don't provide a script or any documentation of how to use can lead to problems where users can't perform the tasks as they wish. The use of scripts or any kind of documentation that can help users to execute tasks or functionalities may be something we will take into consideration in our future work.

2.3.3.3 Captain Memo graph

Another very interesting publication related to the visualization of linked data with a graph layout perspective is CaptianMemo Web Graph [17]. The memo graph is a web application, that provides users the ability to visualize linked data sources as a graph, using a force-directed algorithm. Figure 2.18 shows one simple example of the interface. This method reflects the relative importance of the nodes in the generated graph as it arranges the nodes in a way that the highly connected

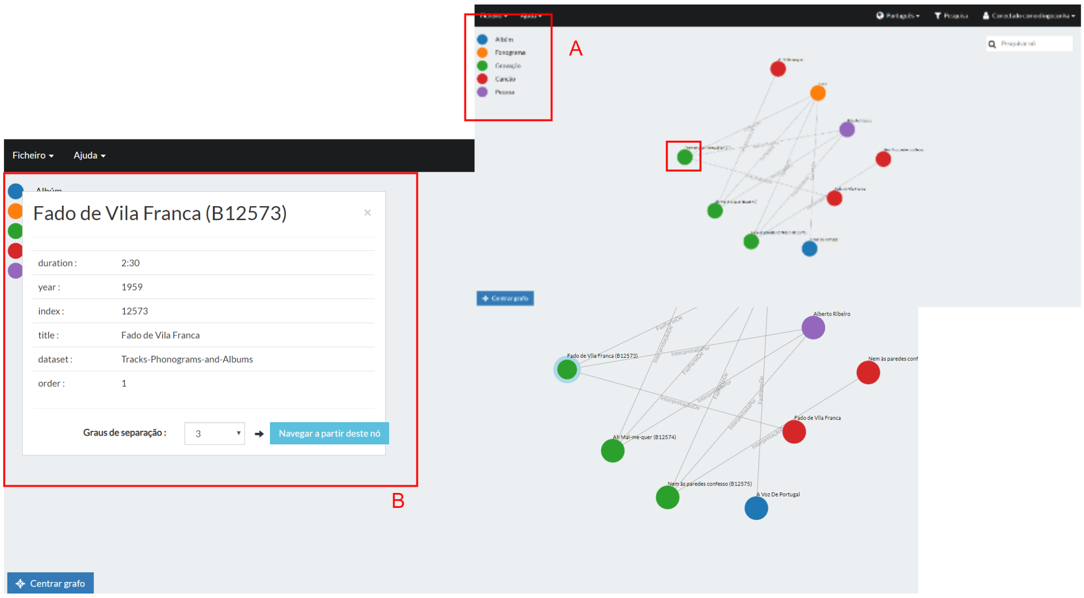


Figure 2.16: Heritamus modal information.

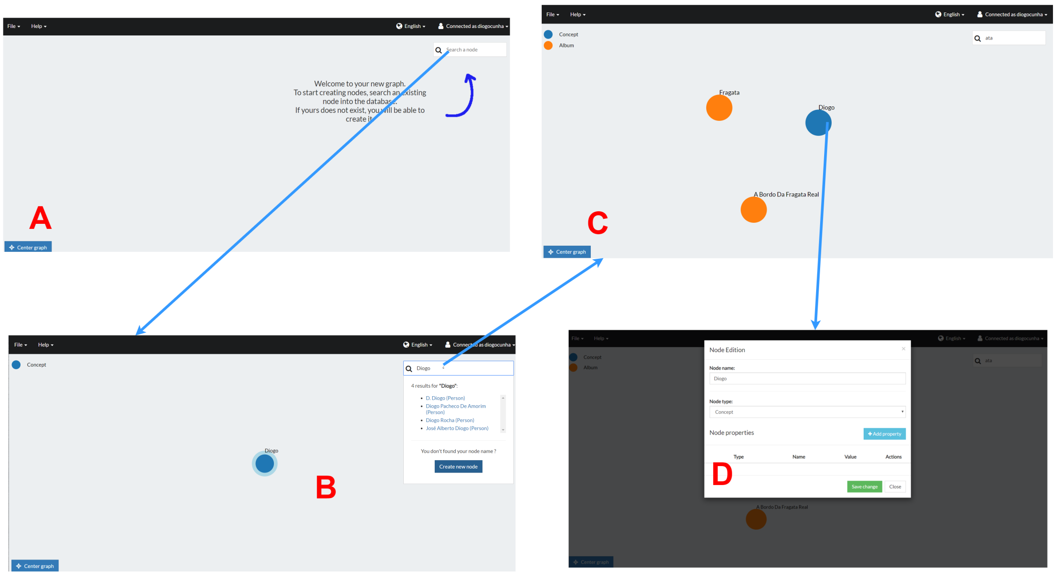


Figure 2.17: Heritamus: workflow for the creation of graph.

Figure 2.18: Captain Memo: workflow for the creation of graph.

ones are placed more to the center of the visualization, while the less connected ones are placed in the periphery.

Using the memo graph, users can have a good understanding of what is more relevant and it increases the legibility of the graph as it tends to avoid the problem of edge crossings. The force-directed visualization authors mention that it can be adapted, since the attraction forces between the different nodes can be modified in favour of a manual repositioning of the elements.

Nodes are identified using both labels and pictures. The picture is automatically extracted from Google if it is not provided by the user. The use of the picture makes nodes distinguishable from each other. Role relations between the related nodes are represented using labelled edges. Memo Graph extracts and visualizes the schema information of the linked data dataset. Class nodes are slightly larger than instance nodes. The authors mention that the layout of the user interface was designed to be as simple as possible and not providing all the information, but only the required to comprehend the graph.

Important aspects related to this web graph engine are, as the authors mention, that the interface is divided into three different parts: the “memo graph viewer“ displaying the visualization of the entire graph, the “memo graph details” listing details about a selected graph node and the “memo search” providing a keyword search option. These are the main features emphasized, but a few more interaction techniques are highlighted to make the navigation easier. Bellow we address some of the features that graph memo used and mentioned and that may be interesting for our future work:

- **Overview**: although Memo Graph is based on an incremental linked data summarization approach, it provides the alternative to show the overview of the whole knowledge base;
- **Filter**: This functionality is provided to let users focus on specific elements in the graph, but the only filter we found was the search;

- **Zoom**: The right clicking on the visualization allows users to zoom-out to analyze the whole structure of the graph or zoom-in to explore certain graph parts;
- **Keyword search**: Memo Graph supports the keyword search task. It highlights the concerned element in the graph and display details about it;
- **Relate**: Memo Graph shows the relationships between the related nodes;
- **Details-on-demand**: To ensure the readability of the visualization, they display details only on demand. We display the datatype properties of the node which is selected. Clicking on a node allows it to be bigger and distinguishable from the other ones;
- **Crossing**: Memo graph tries to avoid crossing edges in the interface to make it look simple and easy to understand;
- **History**: Memo Graph supports undo/redo actions at both macro and micro levels;
- **Animation**: Users can interact with the graph and move the elements around, which results in repositioning of the nodes by animated transitions, generated by the force-directed layout;
- **Non-Connection**: Hide unconnected entities for better understanding of data;

Just as we aim to do in future work, the memo graph has evaluated the usefulness of Memo Graph with real users. This was also a very important aspect to highlight on their publication, as they had compared their system with other already available, RDF Gravity. The RDF(G) is a free, open-source visualization tool Knowledge Information Systems, it supports RDF graph structures and OWL ontologies, but since it isn't a web technology we haven't necessarily investigated in detailed since it was one of our requirements.

The methodologies used for this cooperation had three distinguished phases: Quantitative Study, where tasks were given to users to perform a feature and they registered the timed to compare; Qualitative Feedback, after finishing all the tasks, the participants filled a post-session questionnaire on which they had to rate the performance of the tool that they used; and Free-exploration, by using the method think-aloud, as the users were exploring the website, they were taking notes.

The evaluation of any system and how intuitive it may be for a specific user is as important as the development, and since we will be developing prototypes, the knowledge to understand and know how to proper evaluation is an important aspect, so in the next section, we will address some existing techniques about the methodologies of evaluation for our future work.

2.3.3.4 Conclusions

In this section, we mentioned two web graph engine models that we find interesting to report and describe in terms of use of network graphs. In both tools, we can see some similarities, such as the use of modals or additional information to the node by double click or mouse hover. Also, in the connection of the nodes both systems show what is the type of the relationship between them, which makes the interpretation easier for users. Overall, in terms of visualization, both appear to be easy to understand but it is obvious that it depends on the amount of data. Both web applications, show and provide examples with small graph networks, which makes it easy to find the nodes and links for the connections.

Even with more data displayed, filters can play an important role in graph visualization to make data easier to understand, for example, managing the distance nodes, increasing or decreasing the links, change the node radius, probably changing nodes color for interpretation.

2.4 Analysis and methods of visualization evaluation

2.4.1 User research methods

For the past years, there has been a debate about the scope of user experience, and how it should be defined in terms of evaluation [26]. The field of user experience has a wide range of research methods available, ranging from methods such as lab-based usability studies to those that have been more recently developed, such as unmediated online UX assessments.

The definition of 'User Experience' for test evaluation is defined by a person's perceptions and responses that result from the use and/or anticipated use of a product or a system. This contrasts with the revised definition of usability – the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [24].

So, for both these definitions, it suggests that usability or user experience can be measured during or after the use of a product, and this measure can be very important for every development project phase. Detect a person's "perceptions and responses" in the definition of user experience are similar to the concept of satisfaction in usability. From this perspective, measures of user experience can be encompassed within 4 scope variables which are all connected: attitudinal, behavioral, qualitative and quantitative, as shown in Figure 2.19.

Within all these 4 dimensions, their intersection combines in specific evaluation methods such as eye-tracking, interviews, think-aloud, A/B testing, etc, all techniques in Figure 2.19.

While it's not realistic to use the full set of methods on a given project, nearly all projects would benefit from multiple research methods and from combining insights. Unfortunately, many design teams only use one or two methods that they are familiar with. The key question is what to do when [9]. To better understand when to use which method, it is helpful to view them along with a 3-dimensional framework with the following axes: Qualitative vs. Quantitative, Attitudinal vs. Behavioral, and Context of Use or free to use platform.

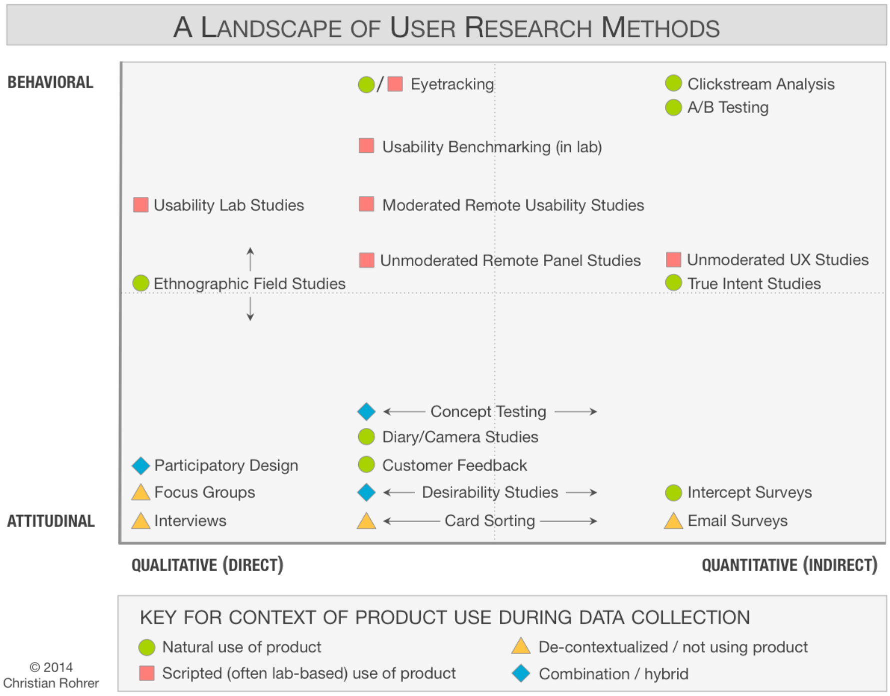


Figure 2.19: When to use which user-experience research methods (from [9]).

2.4.2 The attitudinal and behavioral dimension

When it comes to conducting user research, there are many tried and tested techniques, each with its pros and cons. When we are trying to choose the best technique for any individual project the decision is not considered "black or white". What's right for one project or system, may not be right for another. But essentially, to really get to know your users and create truly user-centered experiences, you need to make sure you not only consider what they say but also what they do and what they pretend to see. So the division into these two categories, attitudinal or behavioral, for the users, may help to understand important aspects.

In terms of attitudinal evaluation, humans are naturally self-conscious and constantly worried about how others perceive us. This desire to be accepted and to fit in can influence what users say, what they think and prevent them from being completely honest. For example, card sorting, an attitudinal method that provides insights about users' mental models of information space and can help determine the best information architecture for your product, application, or website. Surveys measure and categorize attitudes or collect self-reported data that can help track or discover important issues to address. Focus groups tend to be less useful for usability purposes, for a variety of reasons, but provide a top-of-mind view of what people think about a brand or product concept in a group setting [9]. Thus leading researchers to test with a small group of users.

Behavioral evaluation, on the other end, includes methods that focus mostly on the user's behavior seek to understand "what people do" with the product or service in question. Some method from Figure 2.19, for example A/B testing presents changes to a site's design to random samples of site visitors but attempts to hold all else constant, in order to see the effect of different site-design choices on behavior, while eye-tracking seeks to understand how users visually interact with interface designs. Due to the nature of these behavioral research methods, the insights tend to focus on trends identified through analysis of larger numbers of users, in contrast to attitudinal methods which are somewhat constrained by time, sample sizes and available resources.

This distinction can be summed up by contrasting, attitudinal by "what people say" versus behavioral "what people do" (very often the two are quite different). The purpose of attitudinal research is usually to understand or measure people's stated beliefs, which is why attitudinal research is used heavily in marketing departments. And for both dimensions, the interactions with users can be either directly or indirectly. So, in other to fully understand all this stack of information in Figure 2.19, the next section is addressed the following two methods: qualitative and quantitative.

2.4.3 Quantitative and qualitative usability testing

Quantitative and qualitative methods are both complementary types of user research that play important roles in an iterative design cycle [9]. The distinction here is an important one and it can go well beyond the narrow view of qualitative as "open-ended" as in an open-ended survey question. Distinguishing both parameters:

- **Qualitative**: data offers a direct assessment of the usability of a system. Researchers will pay attention and observe users struggle with specific user interface elements and infer which aspects of the design are problematic and which work well. They can always ask participants follow-up questions and change the course of the study to get insights into the specific issue that the participant experiences [24]. Then, based on their own UX knowledge and possibly on observing other participants encounter the same difficulty, researchers will determine whether the respective UI element is indeed poorly designed.

- **Quantitative**: offer an indirect assessment of the usability of a design. They can be based on users' performance on a certain given task, for example: in success rates, the number of errors, or it can reflect participants' perception of usability like satisfaction ratings. Quantitative metrics are simply numbers, and as such, they can be hard to interpret in the absence of a reference point. For example, if 60% of the participants in a study were able to complete a task, is that good or bad? It's hard to say in the absolute. That is why many quantitative studies usually aim not so much to describe the usability of a site, but rather to compare it with a known standard or with the usability of a competitor or a previous design [24].

Comparing both methodologies with the help of Table 2.3, having 4 parameters of comparison may help, such as questions, goals, when to use, which is the outcome and methodologies for each research method. In qualitative research, it is regularly used the word 'why' when users are testing a non-functional prototype; instead, quantitative research normally asks for 'how much' and 'how many' since users are working under a fully completed system.

	Qual Research	Quant Research
Questions	Why?	How many and how much?
Goals	Both formative and summative: inform the design decisions identify usability issues and find solutions for them	Mostly summative: evaluate the usability of an existing site track usability over time compare site with competitors
When it is used	Anytime: during the redesign, or when you have a final working product	When you have a working product (either at the beginning or end of a design cycle)
Outcome	Findings based on the researcher's impressions, interpretations, and prior knowledge	Statistically meaningful results that are likely to be replicated in a different study
Methodology	Few participants Flexible study conditions that can be adjusted according to the team's needs Think-aloud protocol	Many participants Well-defined, strictly controlled study conditions Usually no think-aloud

Table 2.3: Table comparison between quantitative vs. qualitative (from [9]).

Goals are also quite different, qualitative research informs about designs decisions, if is that what users were waiting for or not, while quantitative, for example, tracks the usability over time. When to use, qualitative research is normally used during the process of development, and quantitative is used at terminal stages, so users can compare with other systems. The outcome, qualitative gives researchers impressions and prior modifications, and quantitative provides real statistic numbers for specific functionality. Methodologies for each research method we have the think-aloud protocol, and quantitative A/B testing.

Overall, an advantage of quantitative methodology over the qualitative method is their statistical significance. When quantitative data are presented soundly, they come with a certain protection against randomness: usually, mathematical instruments such as confidence intervals and statistical significance will tell us how likely it is that the data reflect the truth or whether they may be just an effect of random noise — perhaps an artifact of the specific participants that we happened to recruit or of the conditions in which the study was run. While quantitative data can tell us that our design may not be used relative to a reference point, they do not point out what problems users encountered. Even worse, they don't tell us what changes to make in the design to get a better result next time.

2.4.4 Think-Aloud protocol

Our goal is not to test with random users, but instead, to use these methods for the evaluation with professional archivists. We don't pretend to have a huge group of users for this experience, but having proper users and a small group will help for better results.

Thus leading to focus more on ethnographic methodologies (first quadrant in Figure 2.19).

Most ethnographic methods are interactive: they involve dealing with people, users, and stakeholders. A variety of other measures, however, can or cannot require human interaction and can supplement interactive methods of data collection and analysis. These methods require only that the ethnographer keep eyes and ears open using the system, product or application. The ethnographic method is the combination between qualitative methods and attitudinal evaluations, which can be resumed in a few participants, with flexible study conditions that can be adjusted according to the user's needs and with the think-aloud protocol.

The think-aloud protocol is basically a thinking aloud test. Researchers ask the test participants to use the system, while they continuously thinking 'out loud' — that is, simply verbalizing their thoughts as they move through the user interface. In this evaluation technique, is pretended to have users, is intended to have users talking as they navigate the system.

The think-aloud technique is a method that has been studied for years. This entire technique relies on the communication, users need to talk, and so researchers start to propose and see if facilitators (the researchers themselves) when talking would interfere. There are still many aspects of think-aloud usability that deserve serious and systematic research attention. During these interactions, several modifications can be made. For example, facilitators/researchers can or cannot talk during the entire interview, or even participants can only 'think out loud' after completing the tasks.

Many variables can influence this process. Both techniques are called: retrospective think aloud, or concurrent think aloud. Author Henderson et al. [13] mention similar research about this, although it differs in one aspect from concurrent think-aloud protocols: rather than thinking aloud while working, participants initially carry out their tasks working silently and only verbalize their thoughts afterward based on a video recording of their task performance. Theoretically, there are both benefits and drawbacks to using retrospective think-aloud protocols instead of concurrent think-aloud protocols. One benefit involves a possible decrease in reactivity: participants are fully enabled to execute a task in their manner and pace, and are therefore not likely to perform better or worse than usual.

Concurrent thinking aloud, on the other hand, is more prone to reactivity: participants may perform better than usual as a result of a more structured working process, or they can perform worse as a result of their double workload.

Another benefit for a retrospective approach is the recording of working time per task, which is possible in the case of retrospective think-aloud protocols. While users interact with the system there are no stops, but on another hand, this may go beyond the propose of think-aloud protocol, since the requirement to think out loud may be thought to slow down the process of execution in some aspects.

A third advantage would be that participants have the possibility to reflect on their process of using the artifact, which might cause them to highlight higher-level causes for individual usability problems. Overall, the retrospective thinking aloud method may be an alternative to traditional think-aloud tests, since it's probably less difficult for the participants to verbalize their thoughts and understand this method.

Apart from retrospective thinking, concurrent think-aloud protocols also have some drawbacks. Authors Simon and Ericsson [13] talked that a drawback relates to the duration of the participant sessions, which is considerably longer for retrospective think-aloud protocols since the participants not only perform their tasks but also watch these in retrospect. Another important drawback concerns may be the fact that participants may produce biased accounts of the thoughts they had while performing the tasks.

Ericsson and Simon (1993) emphasize that vital information may be lost in the case of retrospective research, which is confirmed by several studies (e.g. Russo et al. 1989, Teague et al. 2001). Very much depends, however, on how participants get to help them recall their thoughts. In the case of retrospective thinking-aloud, participants are immediately exposed to a recording of the entire process they went through, which places the method more or less in an intermediate position between concurrent and retrospective research, and makes it less vulnerable to criticism, which overall is good. So far, there is also some research comparing both methods for the same technology.

Lastly, other authors Dumas and Redish [12], also mention another think-aloud technique called "active intervention". In which basically the test administrator actively probes to get at the participant's mental model or the participant's thought process for how something works. It is basically, providing some help, before users doing the action. Dumas and Redish say that the appropriate technique to use depends on what the goal is, or what the researcher wants, but in the end, they don't offer alternatives to active intervention.

In conclusion, after some deep research about usability, practitioners currently use variations of the think-aloud protocol as their primary way to identify usability problems. The practice of think-aloud protocol varies greatly from project to project, and there are few research articles on which protocol is most effective, but overall, it all comes to the project and goals that are pretended to achieve during the development phase which methodology is the most appropriated.

Previous, we have been able to see how over time a number of tests and adjustments have been made to the think-aloud protocol which becomes somehow flexible and adjustable for each project.

Chapter 3

Interface design and evaluation for user experience interaction

First, we present the platform requirements survey, including use cases, screenshots from the current DigitArq system, and also highlighting some of the requirements and aspects for the new interface. In the second phase, we explain and detail the new prototype proposal, the features that were improved and the new functionalities added. In the third section of this chapter, we present the results from an ethnographic study using the think-aloud protocol. In the last section we present the results of a post-questionnaire regarding the data retrieved from the previous evaluation.

3.1 Requirements from archivists and evaluation of the current application

3.1.1 Evaluation of the DigitArq desktop application

Firstly, in the context of developing solutions for navigating in a linked data system for archival records, we will present the existing system in which the work was based.

The first step was to analyze this existing system that's running under the ISAD(G) model and develop a prototype considering the migration to the CIDOC CRM model, which will allow the system to have new and different ways to interact with linked data.

3.1.1.1 DigitArq desktop software review

For our work, we will detail the main pages, such as view of a record, hierarchical navigation, and it's functionalities. The first impression of DigitArq that users have is presented in Figure 3.1. We have divided the page into 4 parts. First one, on the left side, is where users have access to the ISAD(G) hierarchical model, with all the records listed; Part 2, middle section, is where users will be able to interact directly with the application to visualize and edit the record itself, all functionalities will be listed there; Part 3 is where digital projects are created. These digital projects are the real images printed that represent the record hierarchically in a list of images/photos. Each

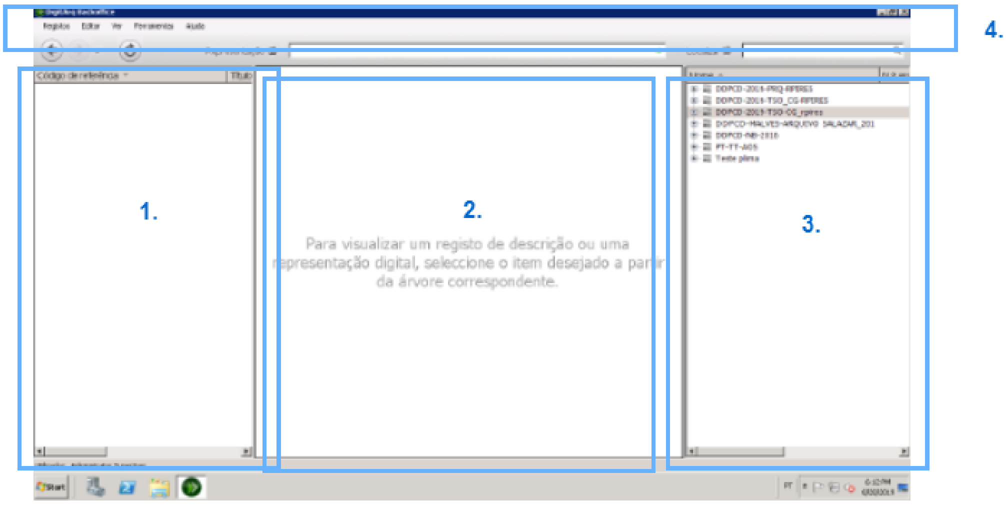


Figure 3.1: First page at DigitArq backoffice.

project can have multiple fonds associated; Part 4, is where users can use the search feature to find records.

Resuming, users have 2 parts of navigation: part 1 to navigate the ISDA(G) hierarchy and part 3 to navigate in the digital records; then part 2 to see the content of the record and part 4 the feature to search for records.

In Figure 3.2 we have the selection of a digital project that we use as an example. The creator name is RPIRES. Then, in the middle section, we're able to see and edit information related to that project such as: general information, properties and who can edit the project. Each project can have multiple records, and inside each record are listed all images and screens of that record.

As we mention before, there are two different ways to edit the document, neither through their reference code (right side, Figure 3.2) or through ISAD(G) description level hierarchy (left side, Figure 3.3). In Figure 3.3 we have the selection of the same digital project as in Figure 3.2, but now by clicking on the left side, we have access to its ISAD(G) content. The following fields and specifications will be important for the development of our prototype.

After the selection of the document, in the middle section 5 separators are presented with different purposes (this can change depending on the ISAD(G) hierarchical entity we choose): Document Identification 'A', User Permissions 'B', Record History 'C', Conversation and Restoration 'D', and Associations and Representations 'E'. In the following images, we will display how the interface looks for each of these 5 components.

In Figure 3.4 we have the selection of option 'A': "Document Identification" and a short overview of the display. This page shows all fields that can be edited in the record.

In Figure 3.5 we have the edit option 'B': "User Permissions". With this option, archivists can change permission to certain records. On the left side, are listed all users available and on the right side, for each user-selected, it is possible to change their permissions. The permissions available for users are: read and publish, read-only, modify only, access only, read, modify and eliminate.

The third edit option 'C' is the "Record History". In Figure 3.6 below users have access to the

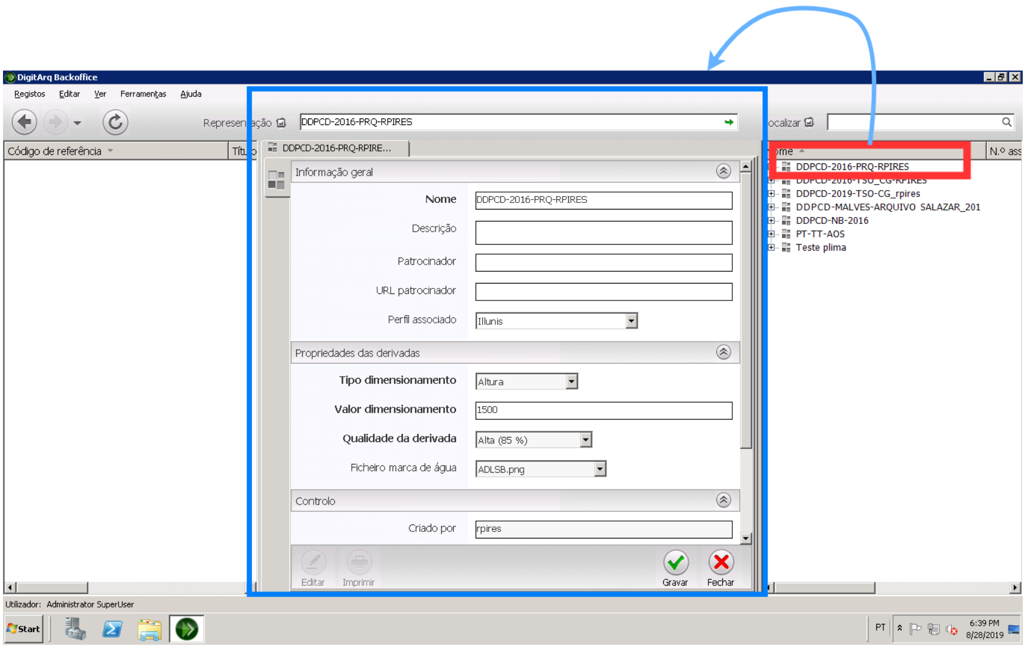


Figure 3.2: Interface for the digital projects .

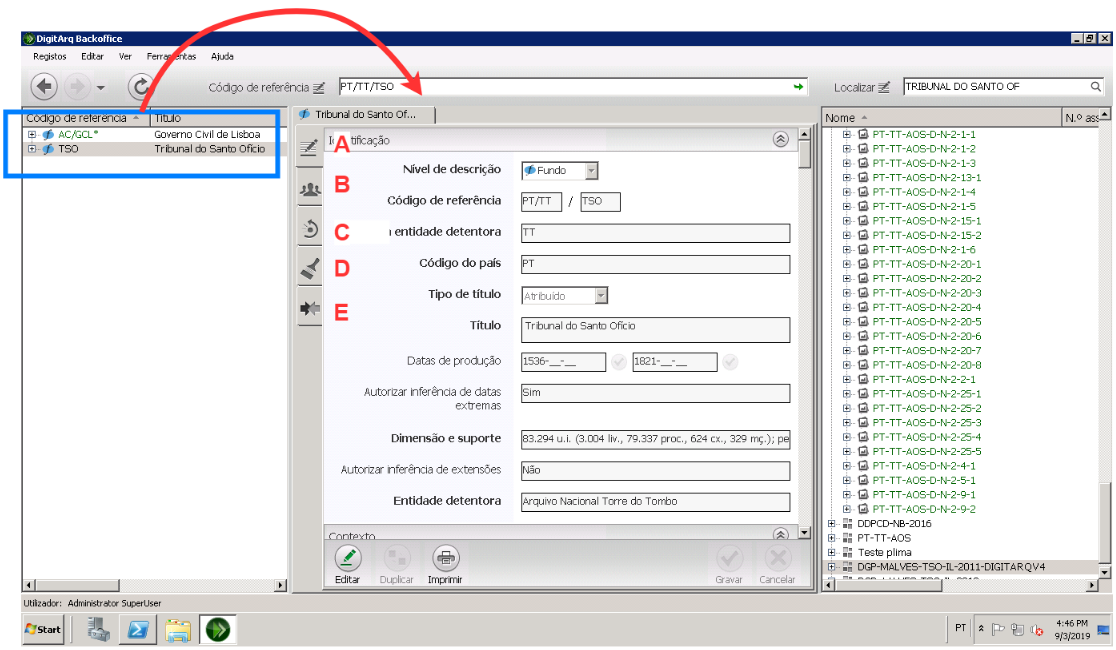


Figure 3.3: Interface for the selection of a fond.

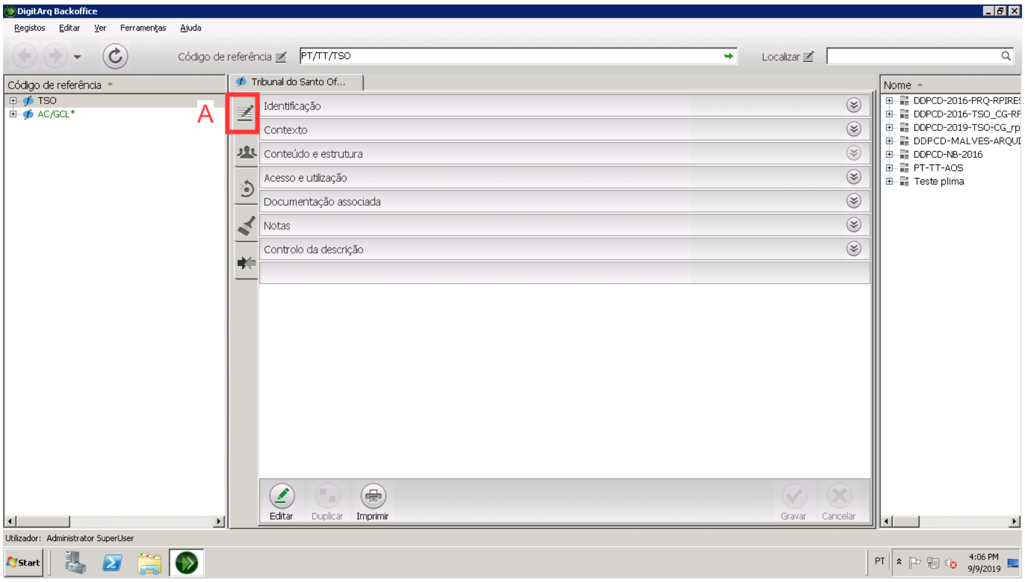


Figure 3.4: (A) Labels inside the record's description.

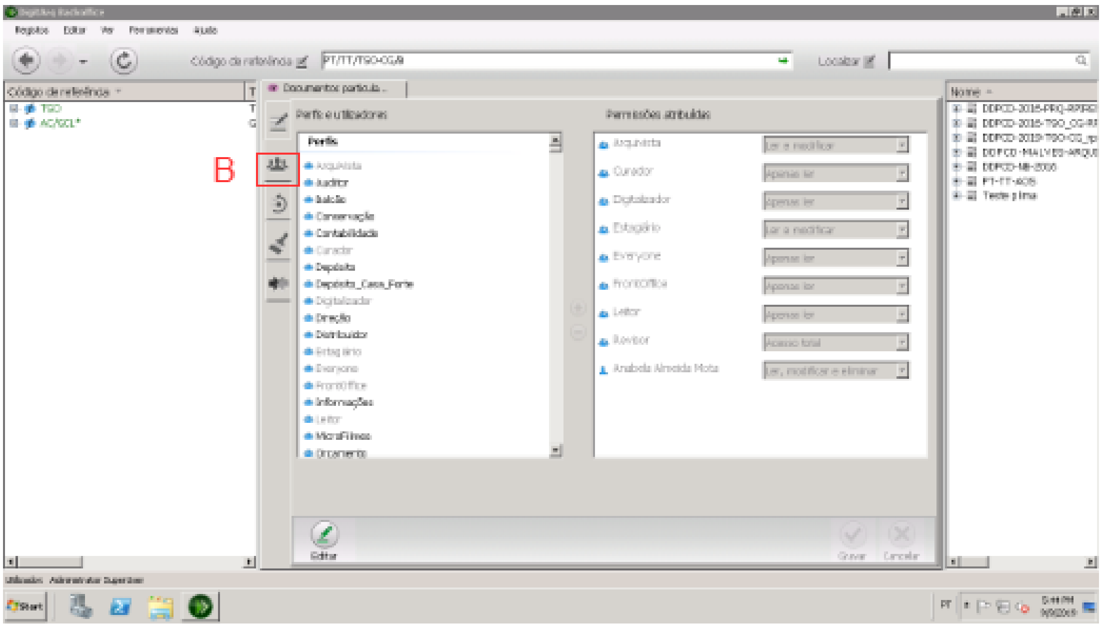


Figure 3.5: (B) Change user permissions.

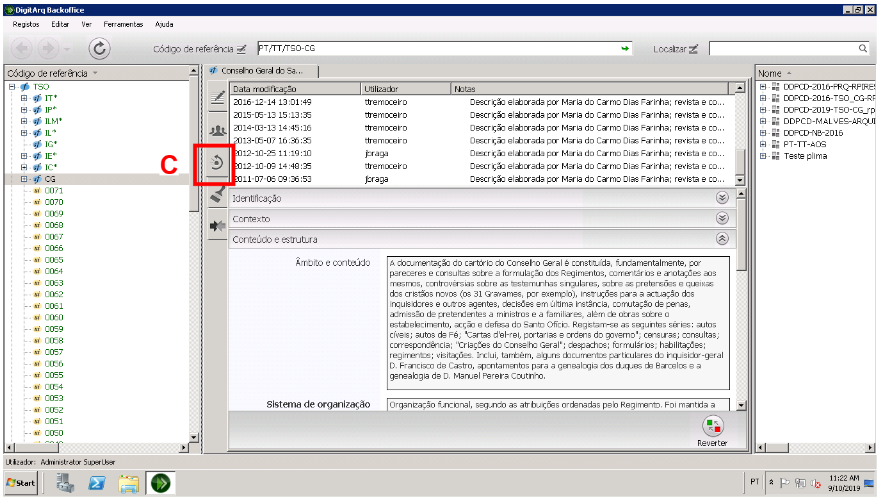


Figure 3.6: (C) Record's history.

ISAD(G) structure to see what changes were made in the record. In the upper section, users can see the latest modification with a short description: Modification Date (e.g.: "2016-04-12 15:44:22"), user (e.g.: "RPIRES") and notes (e.g.: "Descrição mais elaborada na folha 55pag UNI4"). In the section below users can see the record the same way they see in section 'A'.

The next section 'D' in Figure 3.7 is related to the restoration of the record. Users select the button "add restoration" and the program will provide a new page to complete with in information related.

Last section 'E' - Visual representation in Figure 3.8. This section is also important because is where archivists can see the real document digitized. A button "disassociation" is also available to disconnect the image from the document.

3.1.1.2 Conclusions

In conclusion, these are the main pages of the software used currently by archivists. After analyzing the software and talking with the archivists, we've identified particular opportunities for improvement. For example, archivists mention they would like to see the image while they are adding a description or editing the records.

Overall, they mention that adding new features and changing some UI aspects (not completely), in order to make the interaction easier for archivists, would be a great idea. So, in order to better understand the specifications for the prototype, in the next section are listed requirements identified by the archivists during earlier sessions.

3.1.2 Requirements identified by the archivists

Alongside with the development of the EPISA project, meetings with archivists were regularly held throughout the semester in order to comment and document the main requirements for the

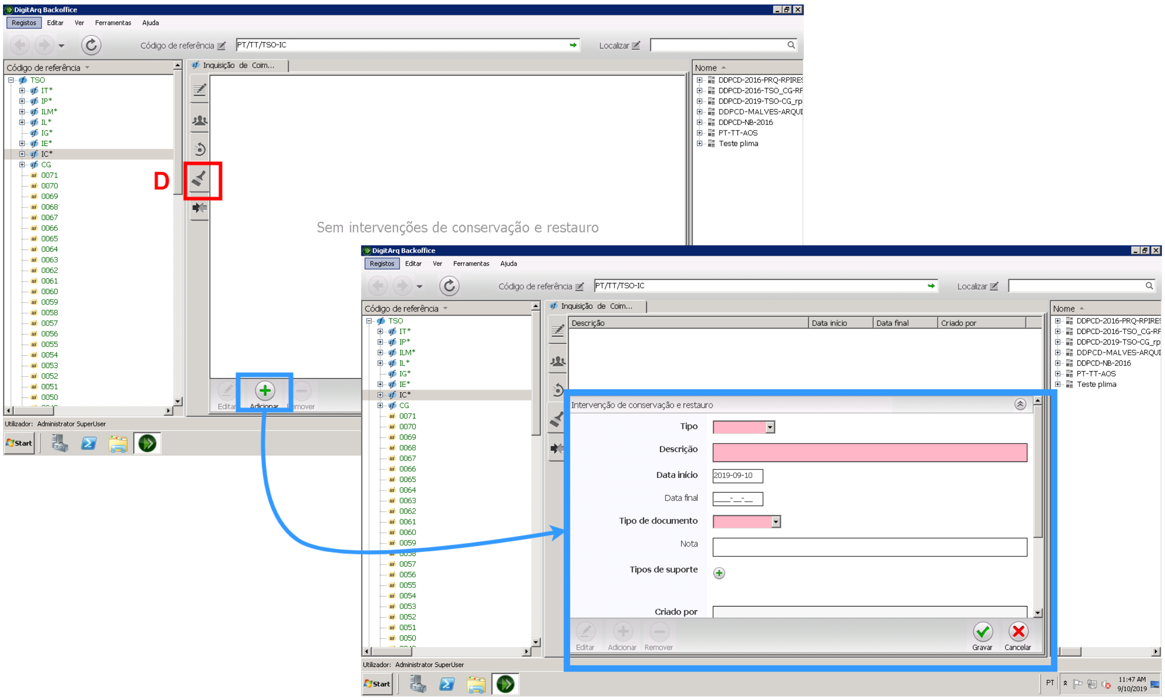


Figure 3.7: (D) Historic of records.

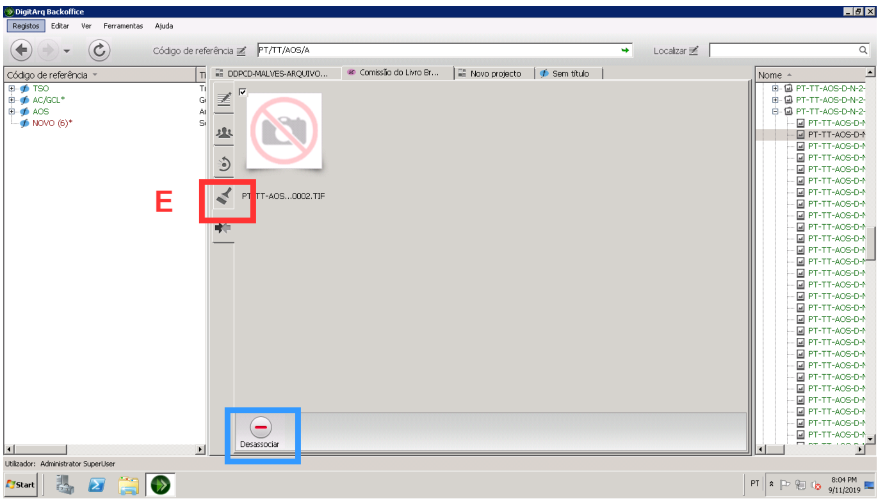


Figure 3.8: (E) Record image representation.

List	Information
A	A new perspective and layout for the system DigitArq+ desktop regarding the new data model for linked data
B	Hierarchical file navigation through a graph view
C	The visualization of related data information while/after navigation search
D	Add a graph layout navigation feature to interact with historical data
E	Create relationships between linked data described by different records
F	Possibility of automating information in records through the system
G	Navigation and manipulation across multiple entities, in both graph and normal visualization
H	The possibility of editing a document, while visualising at the same time an record image
I	Automation of the system for a faster creation of linked records.
J	As an archivists I want to have quick access to my functionality: 'conservador', 'digitalizador' or 'arquivar'

Table 3.1: List of requirements identified by the archivists

development of the new platform regardless of the new data model. The entire list of requirements for the project has much more detailed objectives. In Table 3.1 are important requirements highlighted from the participants and that meet our goals in the dissertation:

Besides all the existing features on the DigitArq desktop system, the change to the new model would enable connection between all types, entities, and actors. Aspects like this were mentioned, as well as the possibility of the manipulation and navigation of data within a graph perspective.

Overall, the most highlighted requirements are based on the interaction and manipulation between linked records and the possibility of graph visualization. In the next section, we will evaluate and emphasize the interaction and visualization techniques within historic records for the DigitArq+ system and later, on Chapter 4 the introduction to the functional prototype for graph interaction.

3.2 Proposed DigitArq+ prototype for linked data in historical records

3.2.1 Main use cases

In the new prototype, the system DigitArq+ aims to apply strategies and visualization techniques from previous studies of linked data and apply them to provide users (archivists) to interact with records and linked entities. Figure 3.9 represents the main use cases and features on the prototype. Normal circles are features already available in the current DigitArq+ system, while bold circles are completely new features developed into the system.

In the next section, and for each use case, we present images and workflows to compare and understand some of these interactions with records and linked data.

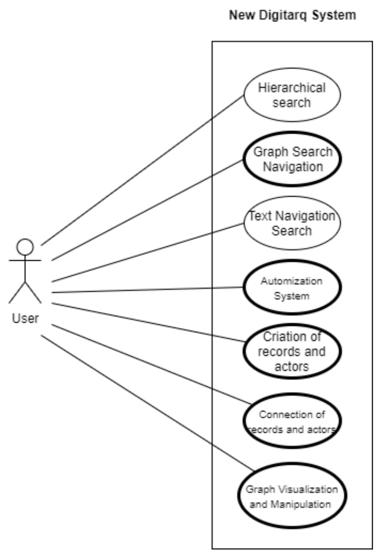


Figure 3.9: System use cases.

3.2.2 Visualization of historical records with DigitArq+.

For future evaluation of these techniques, the study is divided in groups. Next subsections are organized as follows in three distinguish ways of search: hierarchical navigation, text search, and graph search. Then, creation and manipulation of records with linked data.

3.2.2.1 Hierarchical search of historical records

The first feature developed was the hierarchical search in the DigitArq+ system. As mentioned before, archivists used a list for navigation to find the record they want, and in this section, we apply the same functionality but with some particularly different approaches.

In Figure 3.10 is represented the prototype view for the hierarchical navigation on the DigitArq+ system, it consists of the navigation between different ISAD(G) structures that represent a specific record. For the development of this, several aspects were taken into consideration since the current DigitArq system only displays a single line without linked information. Aspects mention in Falcons and Marbels web engines by Berners-Lee [22] include the fact that both systems used different techniques to supply users with more information related to linked data, such as colors or underlined text links.

The main advantage of this visualization in Figure 3.12, comparing with the previous version (blue rectangle), is that it provides much more information for the same selected structure. Not only can users see the reference code, but also other entities related to the record such as actors related, number of actors, locations, and the name itself. The use of colors also makes it easier to understand the type of information. This makes it possible for archivists to navigate through linked data without much effort.

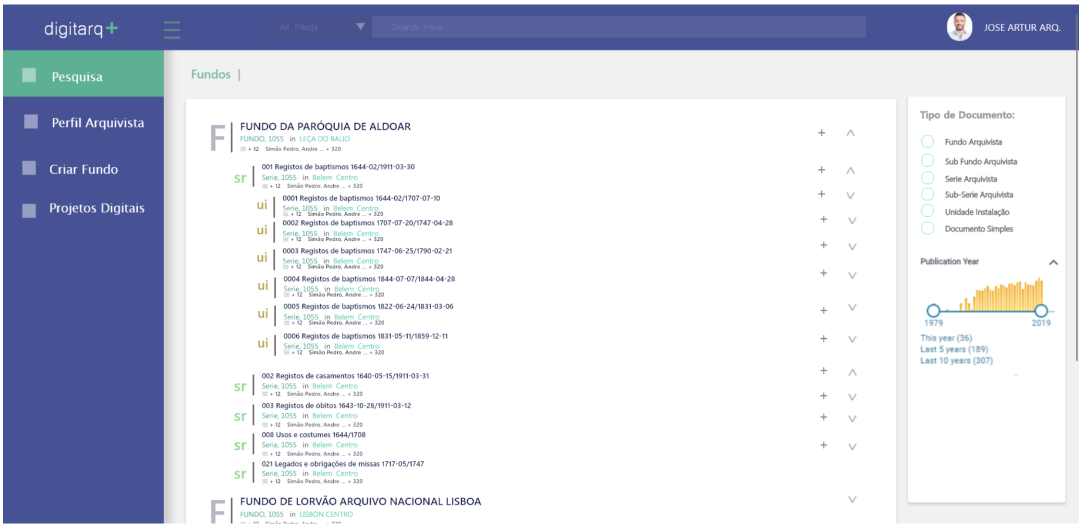


Figure 3.10: Hierarchical search.

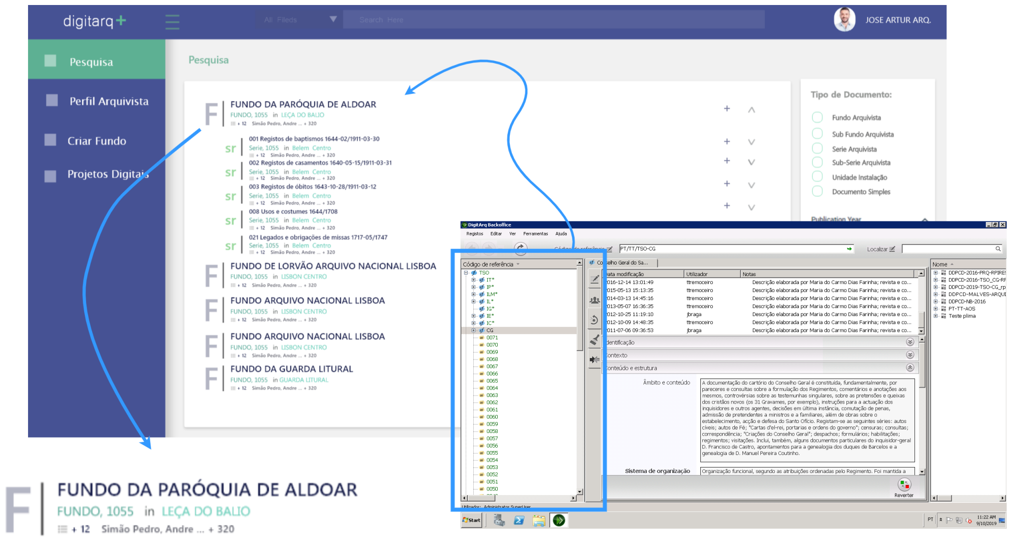


Figure 3.11: Difference between the two hierarchical search approaches on DigitArq+.

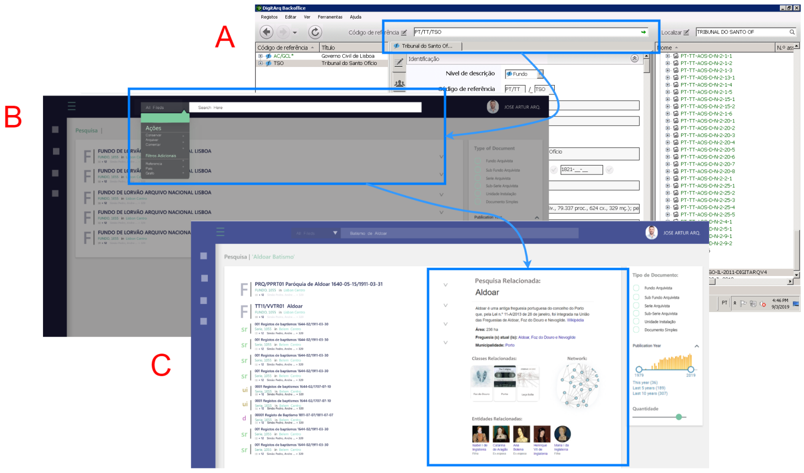


Figure 3.12: Difference between the two search text for records.

3.2.2.2 Text search for historical records

A very common way to search for documents is by using a search bar. The previous DigitArq model already has a similar text search, but it only allows users to search for reference codes instead of keywords.

Below, in Figure 3.11 A, is the search in the actual DigitArq system. B is the search on our prototype, and C is the page where the response is displayed to the request.

The prototype text search in Figure 3.11 matches the archivists' needs, since it provides a keyword search for reference codes, but also provides a more reliable way of quick findings and executable functions as they refer in the requirements. In order to have a faster search, by using a left drop-down beside the search, users can prioritize what they are trying to find, such as persons, cities, organizations, etc. or even the action they want to execute: scanning, conservation, etc. This particularly functionality not also gives users more functionalities to a walk-through, but the result can be much more enriching in terms of records and data shown.

One of the main advantages of using this kind of search within an ontology system rather than a search by reference code, comparing with previous results, is that the result can be much more informative and appealing. On the right side of the page, blue square in Figure 3.11, the prototype is able to show several items related to the search. Archivists commented on the value of this feature, i.e being able to see data that is linked to the keyword searched and presenting actors and organizations that matched the search. On the left side, users see a list of matching records, and, on the right side, linked data that is referred to in the records.

These last two subsections refer to features already available on the DigitArq system, which our goal was to developed different ways of interaction with historical records and linked data. The following functionalities are other ways for data interaction with this historical data.

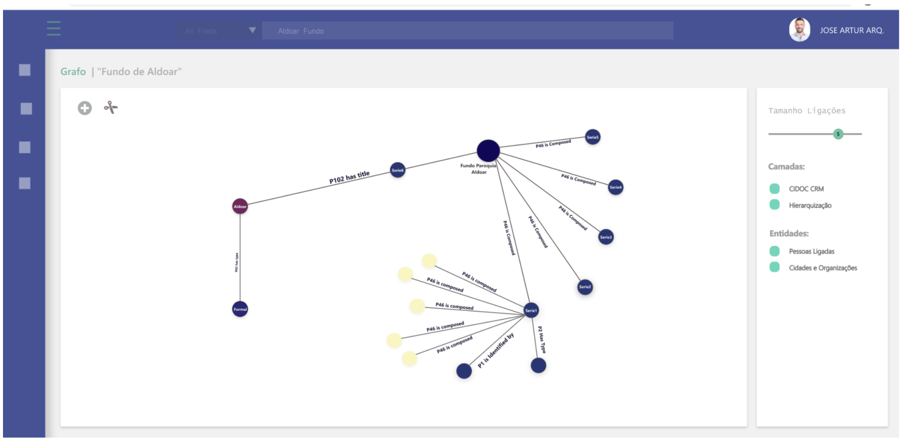


Figure 3.13: Graph visualization of historical records in CIDOC CRM structure.

3.2.2.3 Graph search of historical records

The graph visualization feature was the functionality most expected by the archivists. Users have the ability to navigate throughout the records with a different perspective, plus with the possibility of seeing at the same time entities related to each other.

According to previous studies, the visualization of graph networks, in terms of user experience, is mostly used in the way of a force-directed algorithm [17]. To navigate, each click on a single node, the application expands all the internal nodes that are connected as their CIDOC CRM structure. This way, archivists can see the entire network with the right terminology, in Figure 3.13 is a simple example of a record with its hierarchical and CIDOC CRM terminology.

For navigating the graph, filters were added to make the interaction easier. In Figure 3.14 the prototype provides a simple view without the CIDOC CRM structure, only by showing the real data of the records and eliminating all the CIDOC nodes. In this view, the graph represents the hierarchy from a different perspective.

3.2.2.4 Automation of the creation process for records and linked data

The possibility of automation was another aspect mention by the archivists. Wikidata, for example, provides the auto-complete for the representation of an item, but only when users are writing linked data [14]. What we aim in the new prototype is to provide an automation feature, even before there is any kind of interaction with the records, users can have a faster process for the creation or to link the data.

On the prototype, Figure 3.15, in the right blue rectangle, we present a dedicated section for short and fast functionalities. Regardless of the record, the functionalities can change according to the record. What we aim with further evaluation of the system with archivists is to see if these specific functionalities can actually help when linking and creating records.

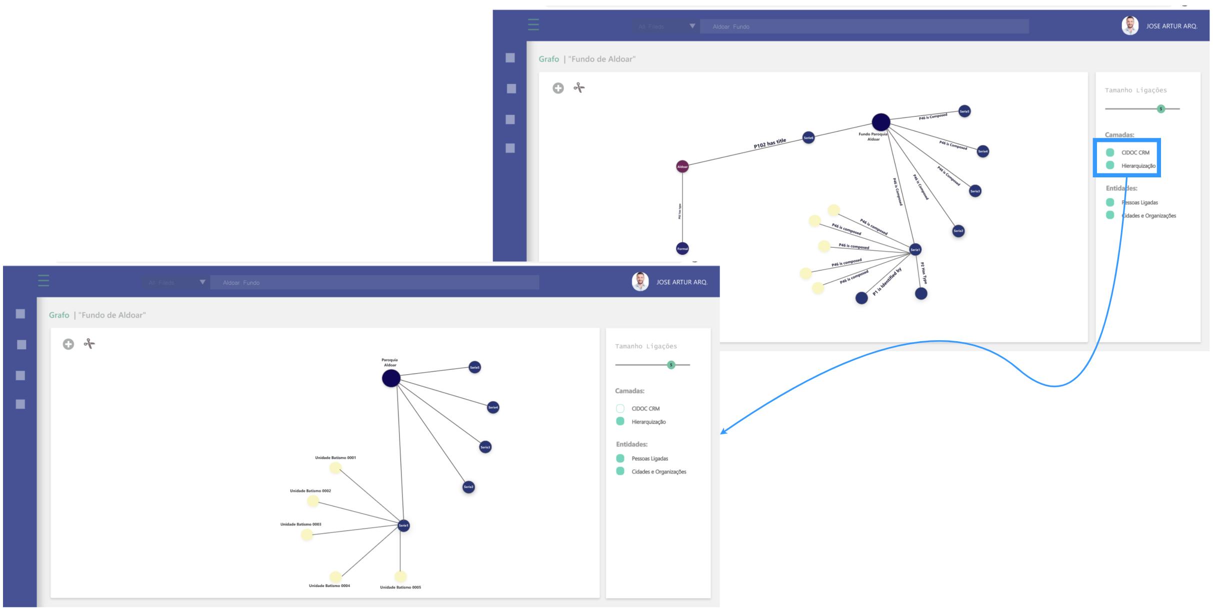


Figure 3.14: Graph visualization of records following a hierarchical structure.

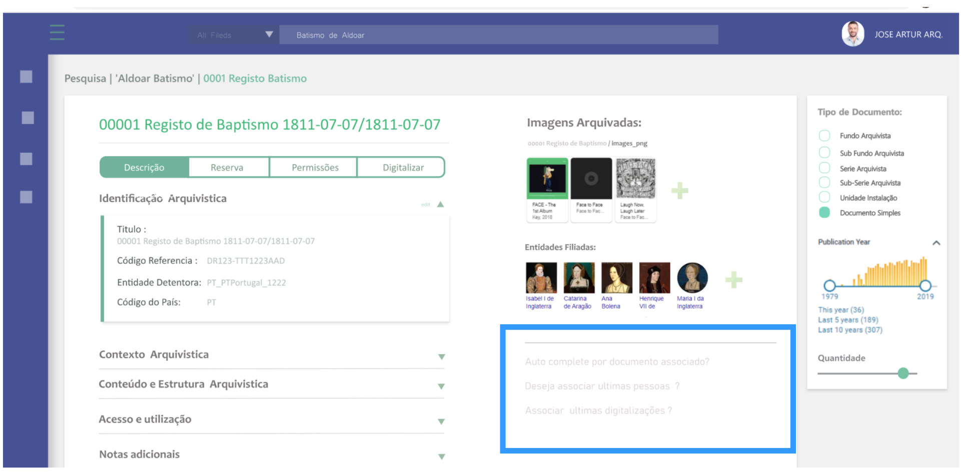


Figure 3.15: Automation section for historical records.

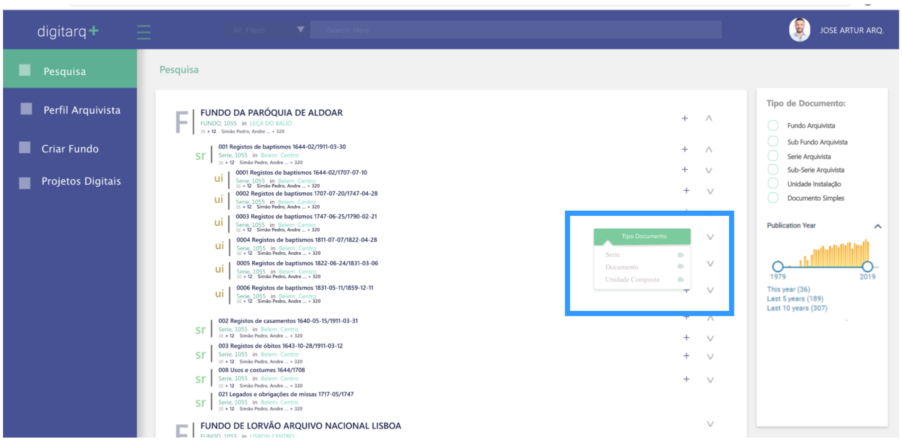


Figure 3.16: Creation of a record associated to his predecessor.

3.2.2.5 Creation of records and entities

In Figure 3.16, a drop-down menu shows the type of record users pretend to create. No single record can be created, only the 'fond'. Any other record is instantaneously connected to the previous one.

For the creation of entities, archivists will only associate entities when finding them on the description of the records, thus leading to make the creation associated within the record itself. This example in Figure 3.17 applies when the entity doesn't exist in the system leading the users to create a new one. On the record's page, users can see a list of entities associated and have access to a 'plus' button to create a new entity. A modal will appear to introduce the information, name, description, type and a list of other entities connected to this entity.

Each record and entity can have multiple associations with entities.

3.2.2.6 Creation of data linked between records and entities

In our previous example, the actor didn't exist. In this example (Figure 3.18), the actor already exists in the ontology system, and the connection between records and entities is made manually by searching for the pretended actor.

3.2.2.7 Manipulation of linked data within a graph view

The same feature as the previous one but within a graph perspective. In Figure 3.19 is presented the creation and connection between historical records and entities related to them. In terms of manipulation, the prototype provides features to the archivists to connect, delete and edit the records and entities.

3.2.3 Diagram and script help for the use of the prototype

For the study and evaluation of interactions with this historical data, we have developed a non-functional prototype that contains several features for our study and the evaluation of the system.

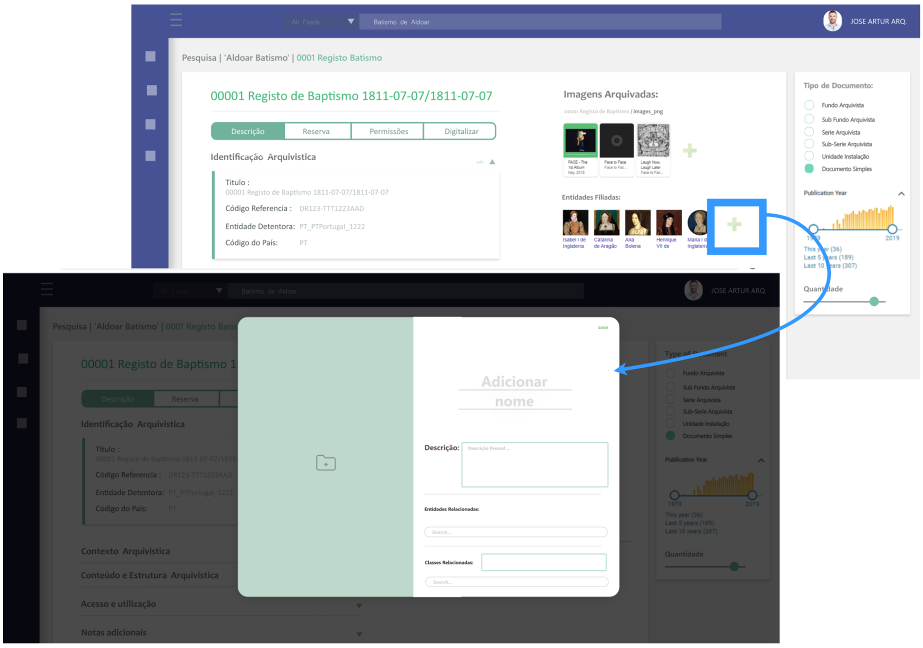


Figure 3.17: Creation of an actor associated to the record.

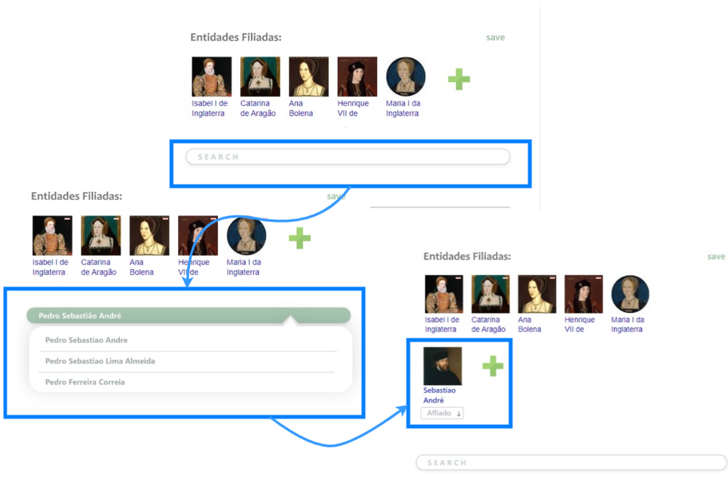


Figure 3.18: Link existing actors to records.

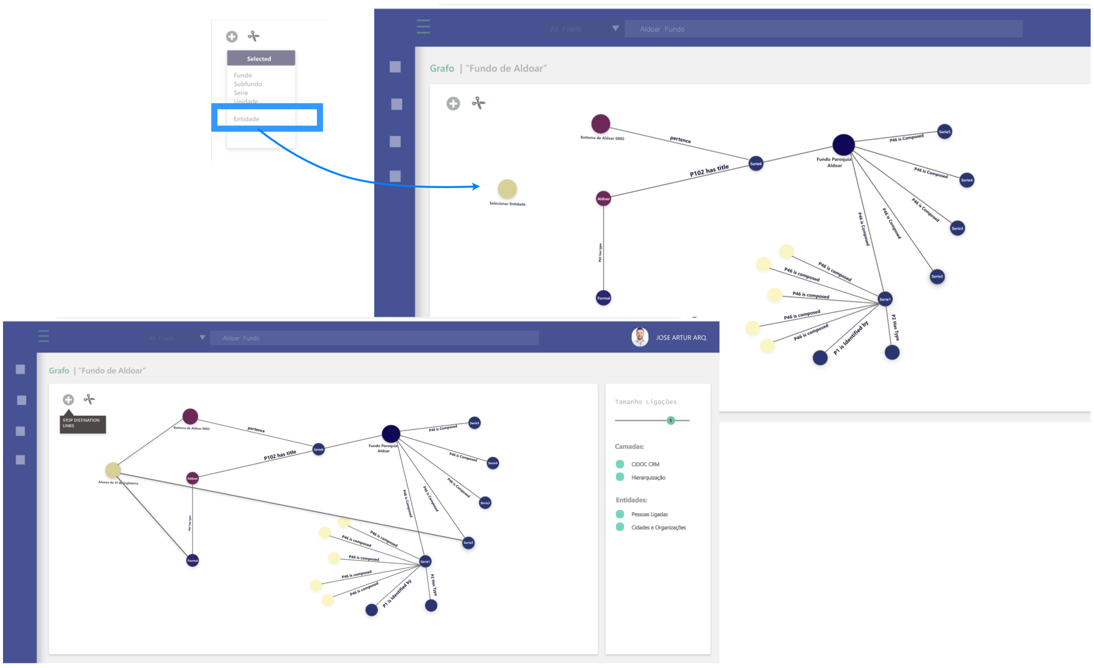


Figure 3.19: Creation of an actor and connection between 3 records.

In order to understand how to use the application, we have developed alongside with the prototype a task script that will help participants to get involved in such a way that participants do not get lost. In Figure 3.20 is the diagram flow that shows the topics the script covers. It is divided into four phases that cover the main use cases in Figure 3.9.

The use of a good task script is essential in this initial phase of development. The importance is to gather the best reliable information when using these usability scripts since users will require careful instructions to successfully navigate the limited functionalities.

Usability testing is a way to peek inside our users' heads and see what they don't like about our prototype or website. The knowledge of what's confusing, frustrating, or stumping the users is the most powerful weapon to turn usability and visualization problems into strengths — thus leading for good and proper evaluation techniques as well. In the next chapter, we will analyze carefully and detail all the studies and meetings with the archivists.

3.3 Experimental procedure for evaluation

This experiment evaluation consists of the execution of four main tasks, as mentioned in Figure 3.20: login, search, features, and creation of historical and linked data. These are the functionalities that participants will explore and test in the DigitArq+ prototype. See also appendix A for the full view of the document. Throughout this chapter, we describe the experiments performed by the participants using the think aloud protocol and create a structured evaluation of all interviews.

All participants will be using the same prototype in the same circumstances. Regardless of the feedback from previous participants, no changes were made after the first interview was completed

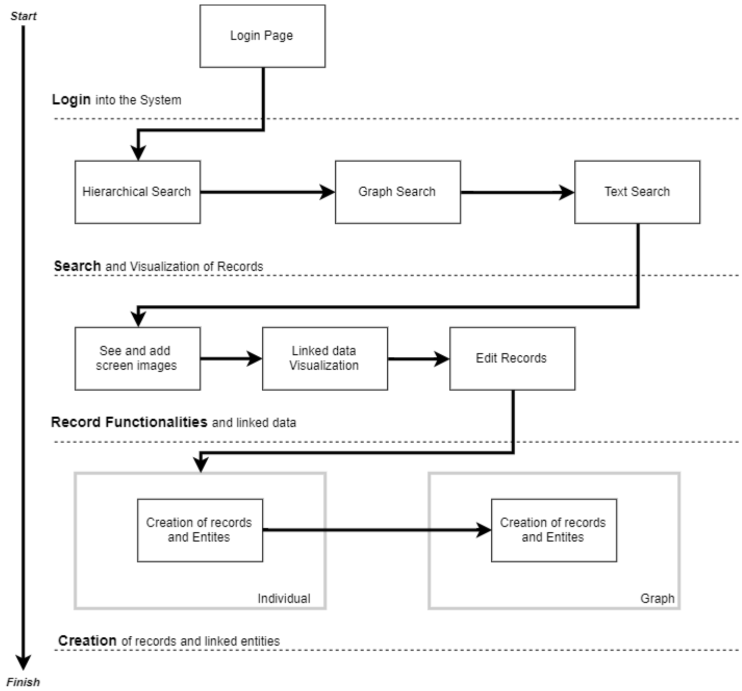


Figure 3.20: Visualization of the script guides for the use of the prototype.

in order to have the most reliable and consistent feedback for evaluation. All participants used the same script to help the navigation in the prototype. After the evaluation of all interviews, a post-interview questionnaire was made according to the use of a prototype in the interventionist sessions.

3.3.1 Participants characteristics for the evaluation

For the choice of participants, the ideal way to test the navigation of the DigitArq+ prototype system would be the selection of users already familiar with the archival context and systems that work with linked data, such specifications meant that no regular users could participate in the interviews.

Considering this, we have selected four persons who work at DGLAB and that we had continuous communication with in the context of project EPISA. In order to evaluate the prototype, we have used the 'think aloud' protocol, which basically let users express their thoughts in real-time about the prototype. In the upcoming subsections, we will now talk about important evaluation metrics that we have collected during all the interviews.

3.3.2 Evaluation study: think-aloud protocol

The evaluation will follow the same topics as our script, mentioned in Figure 3.20. For the remaining study, we will focus only on 3 of the topics: search, record functionalities, and creation. The 'login' layer can be discarded since it doesn't play any important rule for the evaluation of linked data.

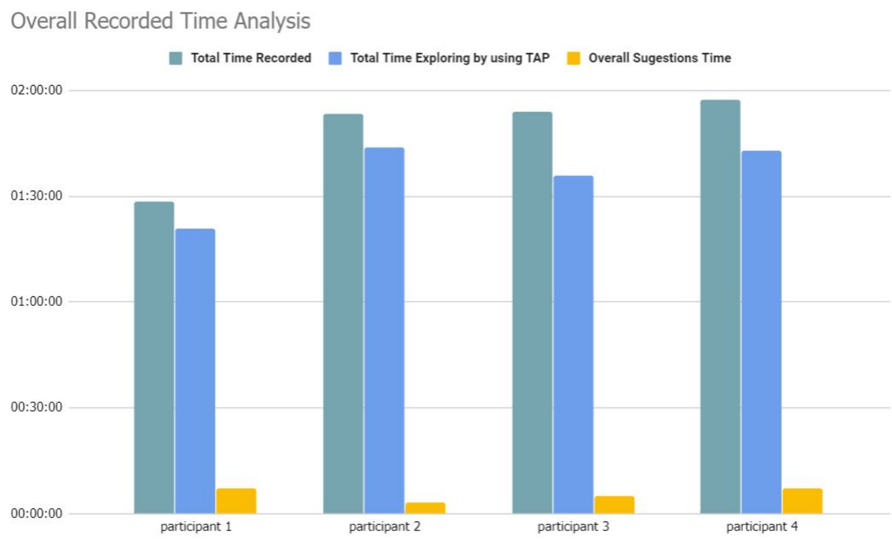


Figure 3.21: Overall times recorded for all four participants.

3.3.2.1 Overall recorded time analysis

All interviews will cover the script topics, but before analyzing each specific action, first in Figure 3.21 is represented as a graphic of all four interactions with the participants during the interviews. Variables TTR (total time recorded) is the entire time since we have started talking with the participants; TTE using TAP (think aloud protocol), is the total time since the interview started and the participants started the navigation in the prototype; OST (overall suggestions time) at the end of each interview we asked participants for a quick overview of the application to see their satisfactions and post-analysis related to the navigation just made.

Before starting the interviews, we have estimated a total time for the participants to follow the script and interact with the prototype of 1h15, in what turned out to be just a bit more extensive. In the end, the average time for each variable is: TTR of 1h48min, TTE of 1h36min, and the OST of 5min30s.

Regardless of the 'login' topic, in Table 3.2 we have the tasks performed by users and all the aspects covered by the participants during the interviews which will be our focus for the next evaluation.

In these evaluation tasks that users will perform from Table 3.2, regardless of being a prototype, we aim to evaluate how well users navigate the system and study the techniques used for the representation and manipulation of historical and linked data. Also identifying which techniques are the most desired as future work.

3.3.3 Metrics for interview evaluation

In order to evaluate and analyze the interviews, it was necessary to arrange some kind of information that could help us to collect and organize topics that were addressed during the navigation of the prototype.

Type	Task	Description
Search	Task 1	Search of historical records using hierarchical search
Search	Task 2	Search of historical records using a graph
Search	Task 3	Search of historical records using text search
Rec. Func.	Task 4	Add Scanned Documents for historical records
Rec. Func.v	Task 5	View record Image related to historical records
Rec. Func.	Task 6	Edit historical documents
Rec. Func.	Task 7	Visualization of linked entities
Rec. Func.	Task 8	Add new linked entities to historical records
Rec. Func.	Task 9	Creation of new linked entities to historical records
Rec. Func.	Task 10	Edit record description
Rec. Func.	Task 11	Archivists functionality of conservation
Creation	Task 12	Create new historical records hierarchical
Creation	Task 13	Create and link new historical records within a graph
Creation	Task 14	Create and link new Entities within a graph

Table 3.2: Tasks performed by users during the interviews.

To better understand this, in Table 3.3 we have listed 5 different groups for the designation of each topic. 'Visualization' can be either positive or negative. We will be looking for positive or negative thoughts about the system. 'Interaction' is the number of times users had a difficulty where to click in the interface, this is a topic more related to design problems. 'Terminology' is the number of problems or doubts users had during the navigation (not visually). Any kind of question they have pronounced, such as "what happens if I click here?" or, "will be possible to list more users?", those kinds of questions that are related to functionalities, and not to the interface. 'Time', is the time spent executing each task of the script. Since we were using the 'concurrent think-aloud aloud' protocol, users were talking at the same time they were executing the script.

To control the time variable, we started counting by the time participants have started to read the script task, and stop when initiating the next one. The total time counting may include all the doubts, problems detected and maybe some future suggestions. We counted the number of suggestions that users refer to in each task. This kind of information may help us understanding if any particular feature caught their attention or not, the more suggestions they bring can mean that they would like to see more out of this feature.

Type	Designation	Meaning
Visualization	PC	Positive feedback about the visualization and functionality
	NC	Negative feedback about the visualization and functionality
Interaction	ND	Number of doubts where to click during navigation
Terminology	NP	Number of problems/questions detected while executing the script
Time	TP	Time performing this task
Suggestions	SU	Additional suggestions

Table 3.3: Topic groups for the evaluation of the interviews.

3.3.4 Data meaning for evaluation

In the following sections, we will display tables with the evaluation notes from the interviews. The direct meaning of numbers from the evaluation is not straight proportional to the number of participants because each participant can have several points for each topic, but this can help us to understand some of the data more easily.

Value zero directly means that no participant has approached the topic. Values from 0 up to 0.75, we can conclude that almost all participants suggested something. Value 1 can mean that all 4 participants agree. Higher values than 1, participants have approached more than once in the following topic.

3.3.5 Search evaluation of data analysis

3.3.5.1 Observation and comprehension of data

The performance of these tasks: 1, 2 and 3, consists of the search of the same record in 3 distinct ways. Analyzing the data in Table 3.4 we can make some conclusions regarding the aspects addressed during the interviews.

Resuming each column of the evaluation, as we can see, the feature that took more time was the hierarchical search, which we were not expecting that since archivist already work with a similar search. We have analyzed the time performed only during the navigation search and it took on average 3.25 minutes, the same as the graph. The rest 7min is the average time spent inside the record's page understating the meaning of data shown. And no positive comments for the hierarchical search, since participants were already familiarized with.

The task with the most positive comments was task 3, the text search, with all participants giving great feedback. This is due to the fact that the result's page provides much more linked information related to the search executed. The fact that records can directly see other records and entities that are directly connected to the search, was well captivating for the participants. Task 2 had more negative comments, most participants mentioned that not being able to centralize the graph was not so friendly, but that's due to the fact we were only using a non-functional prototype, and so participants accepted that problem because after all, the navigation had received great feedback from all participants. The last 2 sections, were as we have expected, the number of problems and suggestions were higher on the text search. Participants were captivated by the functionality, which led many of the participants to come up with a lot of ideas and suggestions. Suggestions addressed some typologies of proper names for the information displayed, but most important suggestions addressed that on the results pages, it would be interesting to have more filters beside the 3 already available, and also, to be able to see not only actors and organizations but also some CIDOC CRM structures.

	PC	NC	ND	NP	SU	TP
Task 1	0	0.75	1.5	0.5	0.75	10.25
Task 2	1	1.5	0.75	0	0.5	3.25
Task 3	2.25	1	0.5	1.5	4	8.25

Table 3.4: Search: evaluated data of the interviews.

3.3.5.2 General comments about the interviews

From the evaluation, we can see participants liked the design of task 2 and task 3. Positive comments were related to the fact users can now navigate on the DigitArq+ system and see linked data and interact with it. Regarding the possibility of combining the text search results with a graph view, no matter how intuitive graph navigation is, participants don't know how well all archivists will perform with it. However, they don't rule out the possibility, since the work developed is not only focused on a current user, but also on future users.

3.3.6 Record functionalities evaluation of data

3.3.6.1 Observation and comprehension of data

In the following evaluation, participants tested features inside the record's page. In this section, instead of analyzing for each topic, we will individually address each task for better comprehension. As we can see in Table 3.5, in several features, participants come with the same evaluations. All rows with ones and zeroes, users have suggested the same, which tells us that both argued identically.

By looking at the chart in Figure 3.23, we can see that most of the time spent on the record's page was during the execution of tasks 7, 8 and 9, the only 3 tasks taking longer than 4.5min. All three tasks are related to the creation and connection of historical data. These tasks caught participants' attention once again. We can see that the longer the users spent navigating and testing the prototype, the more suggestions they provide. Task 7 in the chart was the functionality that participants spent most of the time (TP) and the average of suggestions was also the highest (SU). We found the same pattern in task 8, the second highest time speeding on navigation, and also with more suggestions.

Positive aspects regarding these 3 tasks for the navigation and manipulation, all four participants mention that the features were really useful and that it didn't exist in the actual system. Negative aspects mentioned, task 8 had the highest negative value as well, this result was due the fact users didn't find a good way to select linked data within a list, only selecting the name, Figure 4.11. Problems such as detecting which entity we want to select, if they have the same name, the solution presented did not solve this issue.

Task 4 was another new feature that we aimed to provide, the automation for linked digitized documents. All participants found the possibility of automation interesting ('PC' equal 1), but some negative aspect mentioned was that the digitization scanning was not always done by the archivist, which could lead to some problems. Tasks 5 and 6 is a new feature that can be performed

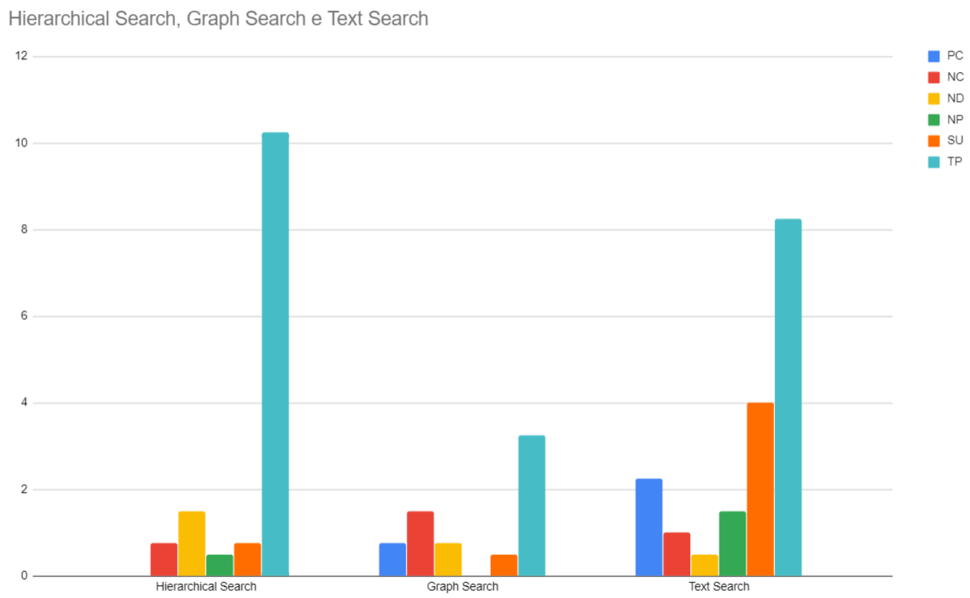


Figure 3.22: Search: chart representation of data from Table 3.4.

as one since one of the requirements from archivists were to edit and see the image at the same time. Most negative comments were related just because they couldn't resize the image, but once again, due to the non-functional prototype, it was not possible, which participants didn't complain about.

For task 10, the possibility of textually connect the record with linked data both 4 participants argued identically, with 'PC' value of one, and saying that the navigation was "pleasant". But in the other hand, some participants notice negative aspects regarding the use of the list to select the pretended entity. Finally, last task 11, as one of their requirements, our goal was to provide a quick method to execute specific tasks. The entire process, at the end participants, found it useful but didn't understand quickly what was the purpose of the search bar, and that explains why the 'ND' value is the second-highest in Figure 3.23.

3.3.6.2 General comments about the interviews

Overall, we can see that for the representation of new features, at least 3 out of 4 participants liked what they saw. One of the goals of the EPISA project is to provide archivists with functionalities to work with linked data, and we noticed on the evaluation study that for every feature related to this topic, users tended to spend more time investigating and thinking out loud about the specific functionality.

In Figure 3.23, we can easily see that for features like tasks 7, 8 and 9, where users test the prototype to manipulate linked data, they spent on average more time on these three than in all others. And the same with suggestions. First and last 2 tasks users, only have performed the script and no suggestions and ideas were recorded, but if we look closer to the center of the Table 3.5, we can see that lots of suggestions were made regarding the navigation and manipulation

	PC	NC	ND	NP	SU	TP
Task 4	1	0.75	1.25	0	0	2.25
Task 5	0.75	0	0	0.25	1	1.5
Task 6	0.5	1.25	0.75	1	0.5	3.5
Task 7	2	1	1.75	3.25	2.75	5.5
Task 8	1.25	1.5	1.5	3.75	1.5	4.75
Task 9	1.75	1	4	2	1.25	4.75
Task 10	1	0.50	2	1	0	2.75
Task 11	0.5	1.25	3.25	1	0	3.25

Table 3.5: Record's functionalities: evaluated data of the interviews.

of historical data. In terms of visualization, we can highlight the use of a list to select linked entities, participants refer that it should have a different approach. Some problems with font size and some intuitive interface handling complaints, due to the fact participants were using a non-functional prototype.

3.3.7 Creation of historical data evaluation data

3.3.7.1 Observation and comprehension of data

In this section, the tasks performed by participants are related to the creation and connection of historical records in two different ways.

By first looking at the chart in Figure 3.24 for these evaluations, we can see the task that took more time on average was the hierarchical search. Since this was the first-time participants were testing the creation feature. During the evaluation we also noticed users had some doubts where executing the tasks, which explains the high value of 'ND'. In task 12, the aim was once again to provide automation of the system with auto-complete of linked data, which most of the participants while executing the script notice who good this feature would be in the new system. Regarding negative comments, all users mention the same due to problems with the interface, wrong archival descriptions, and one participant mentioned that he didn't understand the advantage of the auto-complete.

In the following task 13, users have performed the creation of historical records twice with a total time (TP) of 7.25 on average. Both evaluation methods used the creation of within graph navigation. The difference between these two creations was, one to create a node in which already belonged to a parent node, and the other one to create the node individually. Most of the participants, 0.75 (PC) addressed the creation of nodes pretty intuitive and friendly, the 0.75 of negative comments relate to visualization problems, e.g. the size of the nodes, font and wrong typography used.

Considering the creation of entities in task 13, and by looking at chart in Figure 3.24, we can notice that was the task with less time and points in the table. This may be because the creation of the node actor and a node record is exactly the same, and users were already familiar with the creation after performing the previous task.

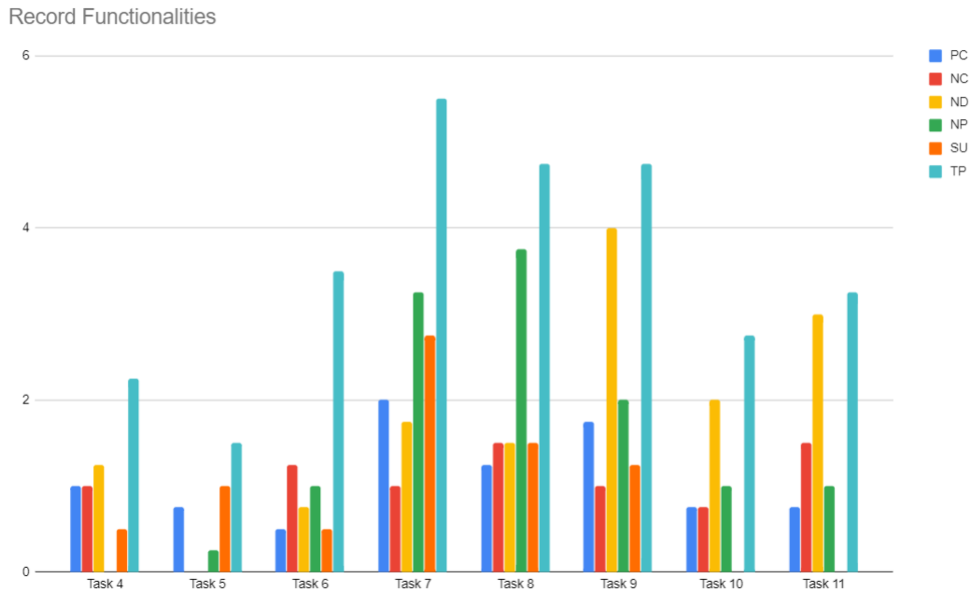


Figure 3.23: Record functionalities: chart representation of data from Table 3.5.

3.3.7.2 General comments about the interviews

Overall, we can conclude that the representation and creation of nodes for historical records, besides being or not intuitive for participants, the average time was actually successful during the interviews. Comparing the time performing the task for creation in Table 3.5, with an average time of 4.75, with the creation with a graph perspective users, have performed in 3.25. And also, mention that during the creation of entities in task 13, participants not only have created a new actor but also connected with three other records. Which in task 9, participants only have associated the actor with only one document. Overall, two participants mention exactly that "the ability to multi-link at the same time from one node is quite good while performing the creation in the graph".

Another interesting aspect detected while evaluating the interviews was that no suggestions were made during the navigation of tasks 13 and 14. Even for task 12, only one participant mentioned three suggestions to handle negative aspects and some new improvements for the prototype. During the document creation, it was mentioned to handle some aspects such as to define which entities are most relevant to appear for a particular type of document (whether it's a person, a geographic location, or an organization) - information that appears on the screen after creating the record.

3.3.8 Observations about the use of a script guide

The current evaluation study was performed under a non-functional prototype. In order to try to collect the best possible information from the participants and to improve the participants' reliable browsing performance while navigating the prototype, we have provided a script to participants

	PC	NC	ND	NP	SU	TP
Task 12	0.75	1	2.5	1.25	0.75	7.75
Task 13	0.75	0.75	0.75	2.5	0	7.25
Task 14	0.5	0	1	0.5	0	3.25

Table 3.6: Creation: evaluated data of the interviews.

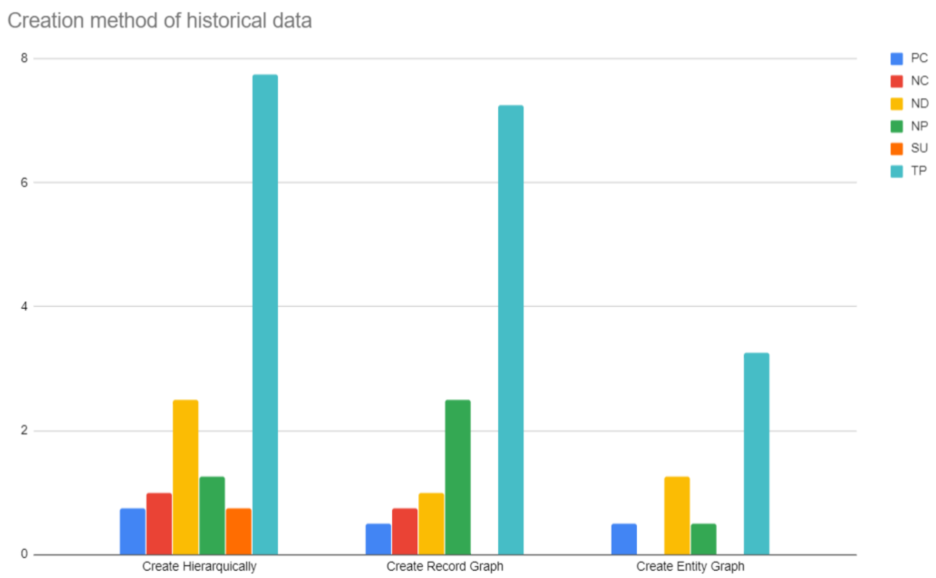


Figure 3.24: Creation: chart representation of data from Table 3.6.

use during the navigation of the prototype. This script has a step-by-step topic to follow, in order to cover all the available functionalities on the prototype, without participants have any trouble where to click.

After the evaluation of the system, we asked participants what they thought about the use of the script while testing the prototype. Most said, that the script definitely helps, they mention that "is a helper and has directions that point to where the person should click - is a guiding tool". If they didn't have the script, the entire evaluation study could take much longer, and it could cause some exhaustion to participants during the interview, and possibly causing some disturbing for later evaluation. In order to control this, some participants also mention the use of a 'timer', to the total time of the interview could be controlled.

Overall, participants mention positive thoughts related to the use of a script while using a non-functional prototype in order to guide for better usability of the system.

3.3.9 Evaluated results and conclusions

In chart Figure 3.25 we have the comparison between the three different types of executions during the interviews: search, record's page, and creation.

As we expected, the highest time performing the evaluation was during the hierarchical navigation. Maybe because it was the first-time participants were interacting with the prototype. Most positive comments were while executing the tasks of the record's page, because this topic has 8 tasks and it may lead margin for participants to come up with more feedback. The highest negative comments were detected during the execution of the search tasks. The highest number of doubts, as well as the number of suggestions, were detected while using the interface on the record's page. This was probably due to the fact that users were performing for the first time the possibility of using linked data, and it also explains the reason why the positive comments are high on the records page.

In the last type of 'creation', we can see that mostly all topics are smaller than the previous ones. In terms of navigation and usability of the prototype, it is an indication that the longer the participants get used in the prototype, things will get more understandable. Or also that the use of a graph for manipulation of historical data, in fact, is intuitive and faster for some cases.

In order to gatherer more information about the navigation experience, after the interviews, we made a questionnaire for participants to answer. In the next section, we will present and discuss the results from this questionnaire.

3.4 Questionnaire evaluation study

In the elaboration of the post-experience questionnaire, we aim to gather information after the sessions using the non-functional DigitArq+ prototype. For the evaluation of the questionnaire, we have used three types of selections: open answer questions, one question for selection of methods, and the rest following the Likert scale to allow the distribution of the answers by levels.

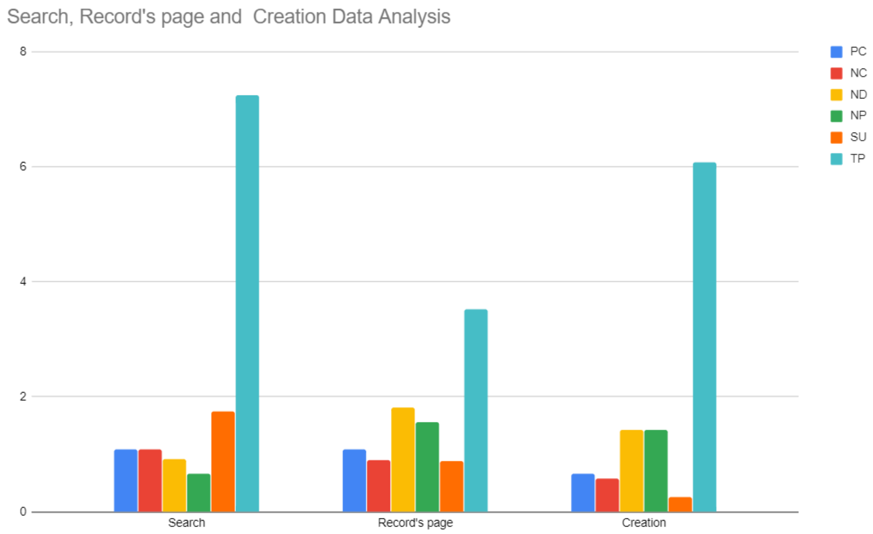


Figure 3.25: Total data comparison between the three evaluation topics.

See also appendix B for the full questionnaire document. We have used the 4-level Likert scale to allow us to obtain the degree of agreement or non-agreement with the imposed statements. Each question on the questionnaire has the following answer options: Strongly Agree, Partially Agree, Partially Disagree and Totally Disagree. The questions from the post-experience are represented in Table 3.7.

3.4.1 Questionnaire post-experience of the prototype evaluation

In question **Q1**, users substantially prefer the text search feature, 4 positive votes, alongside with the hierarchical search, 3 green votes, both features had positive (green) feedback. While in the graph search, half of the participants voted for normal use, and the other half voted for rare use. This result may be because archivists aren't so used to interact with a graph, even if they've mentioned that the use of the graph was pleasant. The results are presented in Figure 3.26.

In **Q2** the results obtained confirm that users want to have two central filters for the navigation: the type of document (fond, subfond, etc) and to be able to see records associated with an institution. Both these filters have the same answer with 3 'very relevant' and one participant voting 'relevant'. The other filter that we can highlight with 4 'relevant' votes is the 'related records' - the filter to see other linked documents. Overall, all others had the same evaluation, regardless of the 'quantity' of the document with the lowest value. All the results are in Figure3.27.

In **Q3** the obtained results show that half of the participants mention that the filters presented in the prototype were enough, the other half details 2 more filters. The ability to filter for events and records with digital representation. This last filter, it was already described in the options of **Q2**, "records with images". The results are presented in Figure 3.28.

In **Q4** we have feedback for one of the most negative comments and doubts during the evaluation of the prototype. In these questions, users record their opinion regarding the selection of

ID	Question
Q1	Knowing that the DigitArq+ system can be used for searching and browsing records, please indicate which method you think has the greatest potential for use
Q2	For filtering the records found, the system may provide different mechanisms. Please indicate the degree of interest for each of the following filter types
Q3	Do you find the above criteria sufficient? If yes, you can move on. If no, you can display other options
Q4	In building links between documents and entities (people, places, etc.) what will be the preferred solution for identifying target entities?
Q5	For the same question, we leave an open answer for suggestions or comments about building links with entities:
Q6	The new DigitArq+ system, following the CIDOC CRM model, will allow each record to be linked with different entities. Order, in order of importance, the entities you consider most relevant in this context:
Q7	For the same question, we leave an open answer for more entity-related suggestions:
Q8	For the following tasks, how important is it to have the fork explicitly visible in the following operations:
Q9	Generally speaking, after using the DigitArq+ prototype, what is your general opinion regarding system navigation and user interface:
Q10	Considering an overview of the prototype you have tried, please indicate other points or comments that you consider important to note:

Table 3.7: Post experiment questionnaire questions.

linked data. The results obtained show that participants prefer navigation in a list to select the user. The same technique used during the analysis of the prototype. This means that the problem may not be in the way of navigation, but how we display the entities to select. The other 2 options, search through an open window, or hierarchy, were well addressed from the participants which could be useful and tested in future work. Result for this question are in Figure 3.29.

In **Q5** the obtained results from these questions, are complementary to **Q4**. Only 3 users suggest different techniques and one participant didn't suggest any other way of selection. The result in Table 3.8 users suggest using a reference code or an identifier (same suggestion as **Q4**), the possibility of having a search combined with different parameters to perform a more effective search. Then, one participant suggests interesting aspects regarding the kind of entity we are linking into the system. If we are connecting a place or a person to a record, participant assumes that a connection through a browser window or even list should be enough. In the case of a smaller aggregation entity such as a document, then greater specificity is warranted and as such, the use of a reference code may be a better solution. All suggestions and results are in Table 3.8.

In **Q6** the obtained result explains many of the suggestions from the participants during the execution of the script in tasks 7, and 9. The results show that 'organizations' is the most relevant entity type. All others had the same feedback, with 3 'very' relevant and 1 relevant. The results of this question are presented in Figure 3.30. In **Q7** the obtained results from these questions, are complementary to **Q6**. Both 2 participants mention the relevance of having entities for events and activities. Another suggestion is the possibility of representation of 'organizations' and

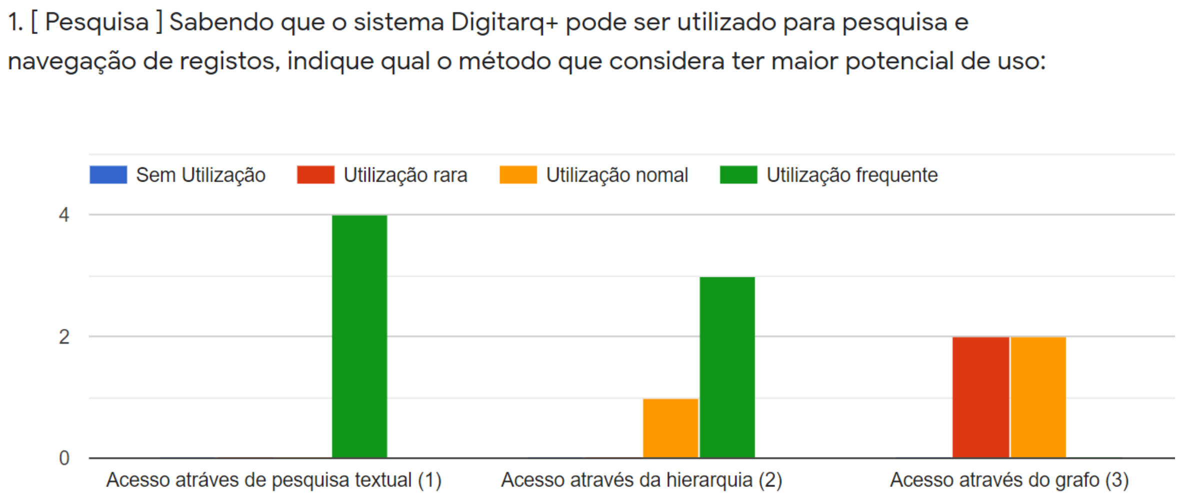


Figure 3.26: Question 1 results on the platform DigitArq+ prototype.

'institutions'. All suggestions mentioned during the questionnaire are in presented in Table 3.9.

In **Q8** the obtained results show that users would only use the graph navigation for navigation and consulting records. These two methods had the same evaluation from the participants. Regarding the use of graph visualization, half participants mention they agree that is partially good for creation, and the other half disagree with the use of graphs for the creation of records. All the results are in Figure 3.31.

In **Q9** the result shows that half of the participants liked the first experience with the prototype for the DigitArq+ system ('above expectations'), while the other half mention that the prototype was 'below expectations'. These results are presented in Figure 3.32.

In **Q10** the result for the last question is a general overview and the participant's expectations about the prototype. The comments go according to question **Q9**, two of the participants didn't mention anything relevant. The other 2 participants mention the exploration of information didn't was according to the model CIDOC CRM, that we haven't explored deep enough all the information we could retrieve, and other participants mention a topic that wasn't explored at all. Historical records cover an average level of granularity. Description rarely goes deep to very low levels such as documents and records, and with the development of the new prototype, they would like to see this particular feature resolved.

3.4.2 Conclusions for the evaluation

In general, all participants have executed all three tasks without leaving behind any step of the script. With the evaluation of the interviews with the think-aloud protocol, we noticed a couple of interesting aspects. Participants suggested more great feedback while performing the search with text, and in second with a graph. We would like to mention that the graph was the most well commented since all the great feedback about the text search was about the representation of the page, rather than the way the search was performed. Overall, the comments highlighted this

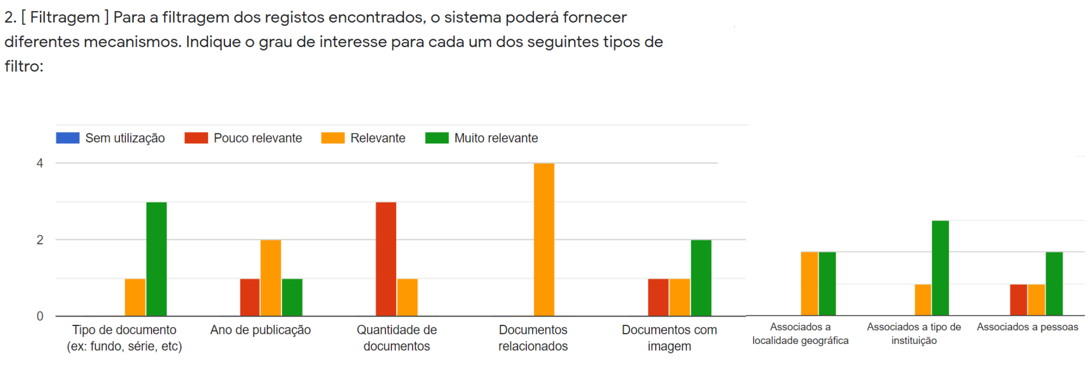


Figure 3.27: Question 2 results on the platform DigitArq+ prototype.

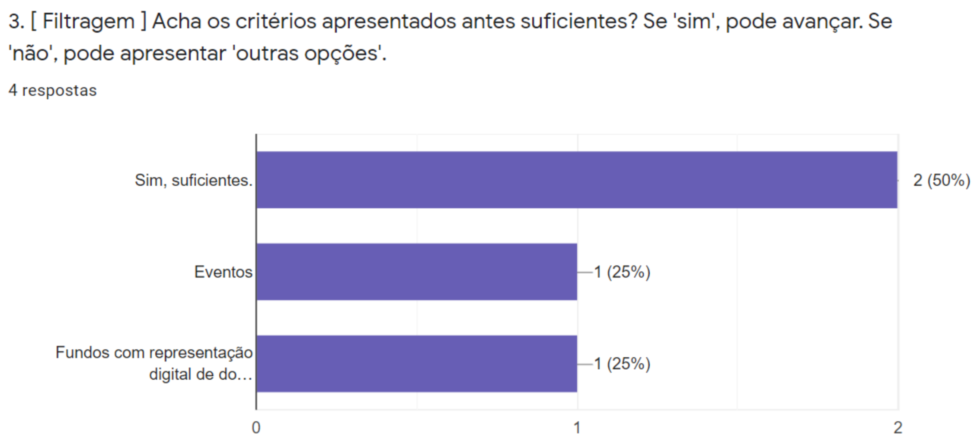


Figure 3.28: Question 3 results on the platform DigitArq+ prototype.

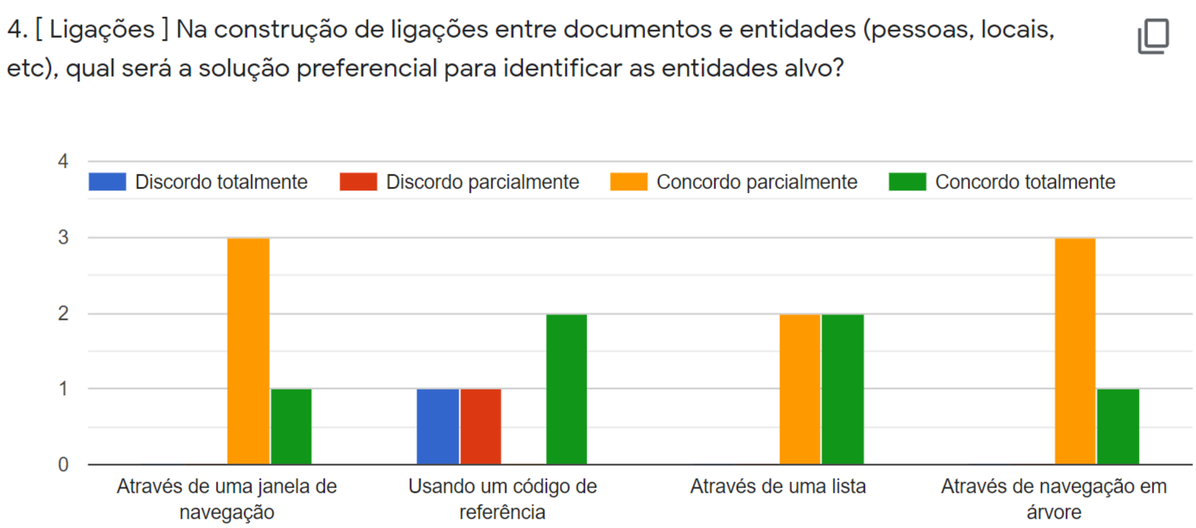


Figure 3.29: Question 4 results on the platform DigitArq+ prototype.

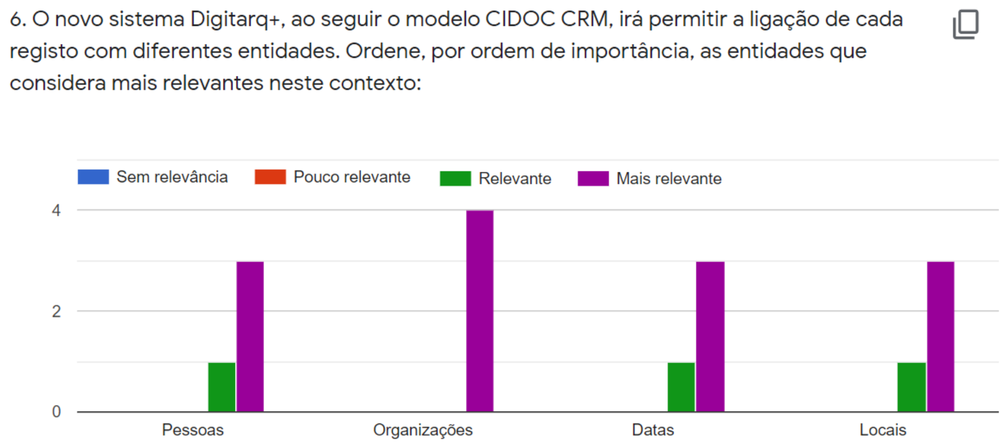


Figure 3.30: Question 6 results on the platform DigitArq+ prototype.

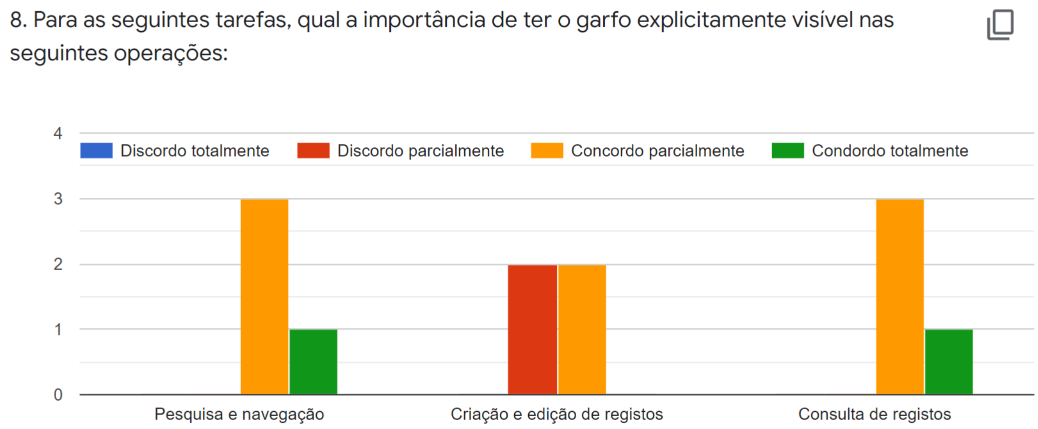


Figure 3.31: Question 8 results on the platform DigitArq+ prototype.

Participant	Answer
P1	The reference code, or other identifier, but automatically
P2	Ideally there should be a possibility to combine search criteria (ex: entities that have been active for a certain period of time, are part of a set or contain certain parts, etc.)
P3	I think this question may depend on the level of granularity or scope of the entity. If you are relating a place or person to a background you may assume a connection through a browser window or even list (e.g. backgrounds); In the case of a smaller aggregation entity such as document then greater specificity is warranted and as such the reference code may be a better solution.

Table 3.8: Q5: List of answers to question 5.

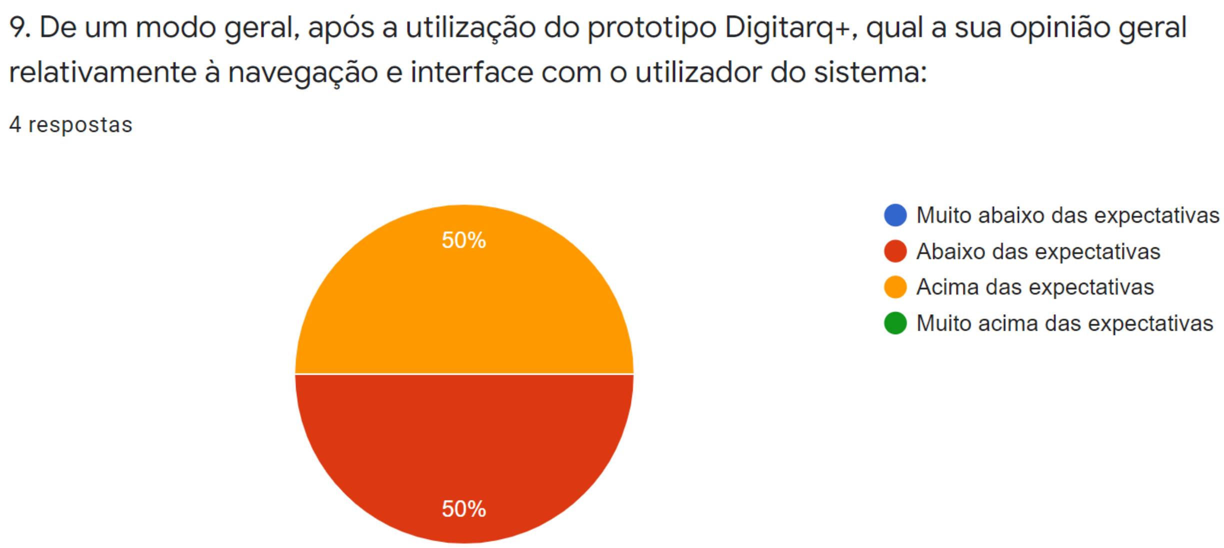


Figure 3.32: Question 9 results on the platform DigitArq+ prototype.

Participant	Answer
P1	events, activities
P2	Here again the reference to Events (and their subclasses) that will also be represented in the new model is missing.
P3	People are understood, in the sense of producers of personal records. The idea of archiving corresponding to an organized documentary set corresponding to a universe of relationships that are usually the parts of the whole (organic and functions). For this reason, Documentation Services and the functions that fulfill the reality of these services
P4	Linking the registration for a document or document aggregation with the registration of another document or related document aggregation (Related Description Units in ISAD (G)). Linking the described units to external resources (described on platforms other than DigitArq+, including those created and managed by communities of practice other than archiving).

Table 3.9: Q7: List of answers to question 7.

Participant	Answer
P1	The position of the observer against the data structure (navigability)
P2	The prototype seemed too "stuck" to the concepts of archival description according to the traditionally used norms and could go further in exploring the information modeled according to CIDOC-CRM.
P3	Note that historical archival description work generally spans an average level of granularity - packet, book, cxa, installation unit - that is, description rarely drops to very low granularity levels such as document and record , given the extent of the fonds (documentary masses often quantified in linear meters and in some cases reduced to Km). Exception to this rule only applies to projects or other rare cases. Ideally, this trend can be changed and technology can contribute precisely to foster a more granular approach. I am focusing on this to draw attention to the fact that issues of detail relations do not correspond to the common practice of archival description, although I also think that there may be an advantage in improving.
P4	Nothing to say

Table 3.10: Q10: List of answers to question 10.

feature, i.e. all four participants mentioned this navigation intuitive and "nice". As a result, we can see that the time performed in the graph comparing to others, was the smallest (chart Figure 3.22). Emphasizing this conclusion more.

Considering the record's functionalities, users have highlighted the possibility of seeing linked data within this archival context. With tasks 7, 8 and 9, it helped to emphasize precisely this conclusion. When comparing the connection of entities in both perspectives, the creation of links in the record's page (task 8), on average users performed in 6 minutes, while the creation of entities and linking the data with several other archives, on a graph perspective users took 3 min. Even though the evaluation uses a non-functional prototype, we can conclude that the use of a graph may truly help archivists in some specific scenarios when linking specific data and records.

The second part of our conclusions, with the questionnaire, were slightly different. Overall, users have mentioned that the search method they would use the most is the text search, and the graph, not so often, but if they use it, they suggest the graph is better for the search and consultation of archives rather than for creation. Comparing these results with the interviews, we think that part of this inconsistency is because archivists aren't used to graph functionalities.

Overall, half of the participants liked the use of the prototype, and the other half suggested that we could have gone more in-depth and explore more about the CIDOC CRM model for representation.

Chapter 4

Implementation of a prototype for graph-based interaction

In this chapter we describe the development of a prototype to manipulate and interact with real historical data using a graph-based approach. For the development of this prototype, we have taken into consideration the feedback from the participants during the interviews, which helps us to decide which features, filters, and important parameters should be emphasized.

In the following subsections we'll describe the technologies used for the development of the graph interface and also provide a workflow diagram of how to use the prototype and all the features and filters implemented.

4.1 Implementation details and programming engines

For the development of the graph, we have used HTML, CSS and, from all the technologies presented in Table 2.2, we have chosen D3 as our framework for graph development. We have selected HTML and CSS because of the comfort with the technologies and the ease of use.

About D3, since it was our first time developing with this technology, one very important reason to choose was the documentation available online and all the features already available that help us to accomplish the desired result. The possibility of D3 having more than +100 different layouts with smooth transitions and interaction was also an advantageous aspect. We used the library d3.layout.force version v3.0. The use of SVG by D3js also provides us with a good reliable way of creating interfaces.

4.1.1 Page layers

The way the web graph application is structured is presented in Figure 4.1. Below the image we have represented a user interacting with the system. The web page is composed of two HTML pages, the first where users can see the entire graph (graph.html), and another page to see an individual record (document.html). Both pages are connected to the same CSS file, calling the Bootstrap framework for styling the interface for better interaction, and the same JavaScript file to

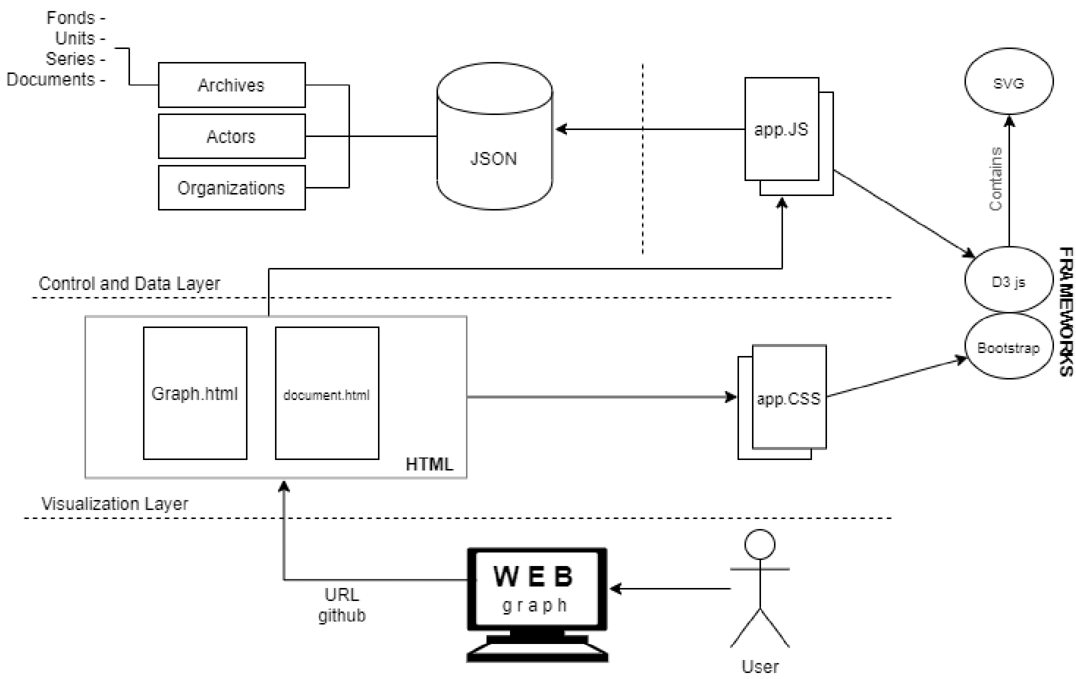


Figure 4.1: Workflow diagram for the back end.

control the workflow during the manipulation. The last layer is divided into two important parts, the JavaScript file that controls all the features available: navigate in the graph, use filter options, create new nodes, create links, edit graph nodes, see the historical record page, and interacts with our JSON data.

4.1.2 Historical data

For the visualization of the historical data, and in order to create the graph, the app.js file extracts all archives from the JSON file. The data used in the prototype was extracted manually from the DGLAB website. We collected several records from the "Fundo de Aldoar". JSON data was structured as shown on the right side of Figure 4.2. We haven't used all fields for the representation of the records, only some important fields to identify each node and visualize the record according to its type. For the representation of the links, we have a different structure. On the right side of Figure 4.2 we use four fields to identify the links: 'source' is associated with the parent node, 'target' is the destination child, 'left' and 'right' variables are for representation purposes to draw the direction of the arrow.

With these two representations, we are able to draw the entire graph and make it possible for users to manipulate and interact directly with historical data within a graph view. Finally, the data represented in Figure 4.1 can be organized in three topics. As mention in previous sections, variables can be more, but for this prototype, we have only implemented these: 'actors' are the representation of real entities; 'organizations' are institutions or real places; and 'archives' can be defined as 'fond', 'unit', 'series' and 'document'.

```
{
    "id": "8811233HRT1233",
    "code":"PT/TT/DME",
    "name":"DIÁRIO DA MANHÃ E ÉPOCA",
    "reflexive": "false" ,
    "year": "1955",
    "type": "FUNDO",
    "description": "O Diário da Manhã:
}
```
Record Node

```
{
    "source": "nodes[0]",
    "target": "nodes[1]",
    "left": "false",
    "right": "true"
}
```
Link Edges

Figure 4.2: JSON structure.

In the next sections, we will describe all functionalities and the interfaces developed for the manipulation of historical data.

4.2 Interfaces of the functional prototype

4.2.1 Web-graph page interaction

The DigitArq+ web graph prototype [1] is live and accessible online. In Figure 4.3 is presented the first impression users will have with the interface.

The interface is divided into three different parts, 'A' is the navigation bar where users can select to see data in the graph, or in a table, or use text search; Part 'B' is the main navigation and visualization of the network graph, and 'C' on the right side are the filters to control the graph.

4.2.2 Filter techniques

In section 'C' we have five different filters listed and available to use in the graph. The first filter to interact with the graph is the color change. Users can change the color of nodes to help to identify

[1] Web Graph Application

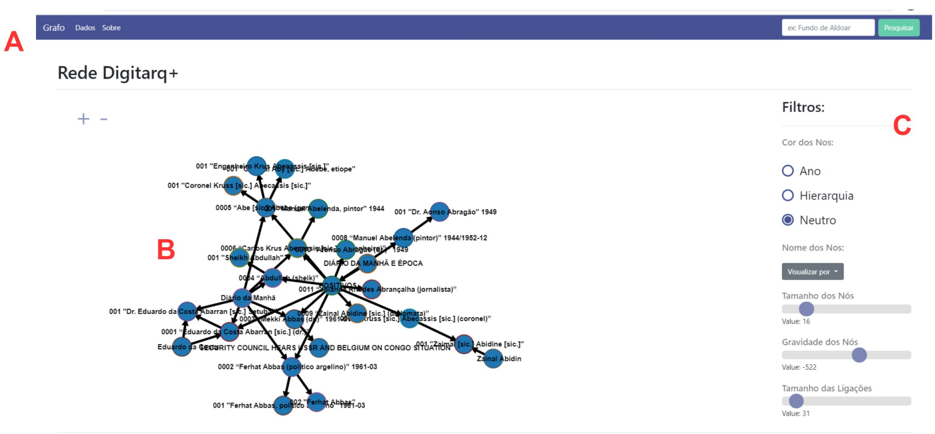


Figure 4.3: First impression of the web-graph engine.

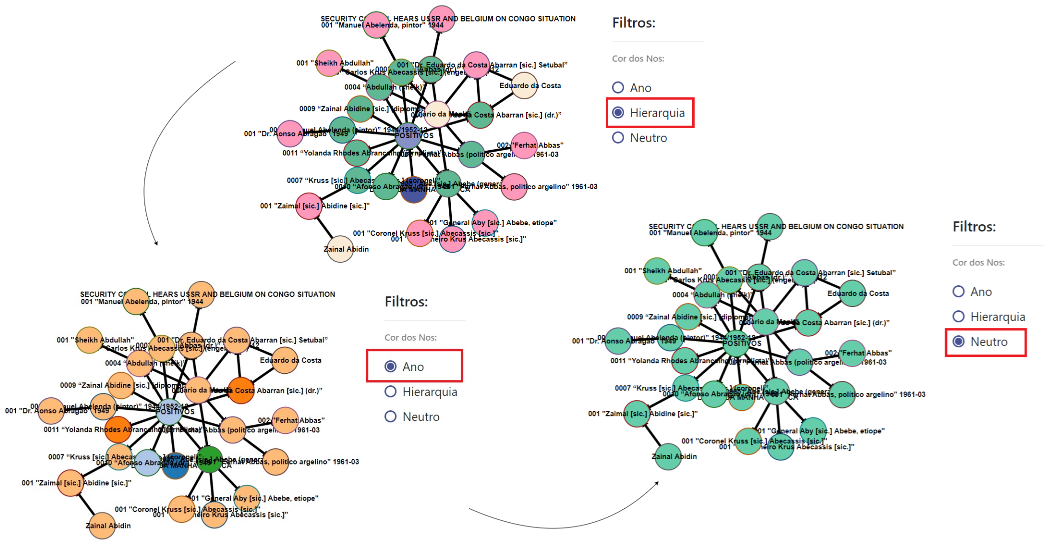


Figure 4.4: Filter for coloring nodes.

the nodes. In Figure 4.4 is represented how the graph looks in these three different perspectives. Users can see the nodes painted with the same color by the age of the record and the entity that was created. This helps users to get a quick overview of which nodes are related to the same year or not. The other option, users can color the nodes with the same hierarchy value following the ISAD(G) structure. Purple colors are fonds and series, green is used for units, pink is used for documents, and light brown are the representation of all entities (actors, organization, etc). The last color is simply to unify and keep all nodes as equal.

Next filter, users can change the node's name in Figure 4.5 for: the name of the document, type, or ID. The last option 'none' is to clear the interface and present no text on the graph. To make it easier for navigation. In this case, with 'mouseover' on the right side of the interface, users have access to a short description of the record.

The last three filters are to change the dimensions of the graph and provide users the ability to adjusts the nodes regarding the amount of data shown. In Figure 4.6, with option 'A' users can change directly the size of the nodes. Options 'B' changes the gravity of nodes - 'gravity' is a variable of D3 that adjusts the repulse when nodes are getting closer to each other. The last filter, 'C', is to change the length of the links. Overall, users can adjust the graph size and perspective

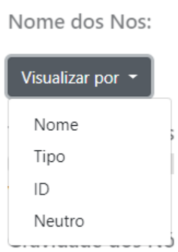


Figure 4.5: Change name of the node on the interface.

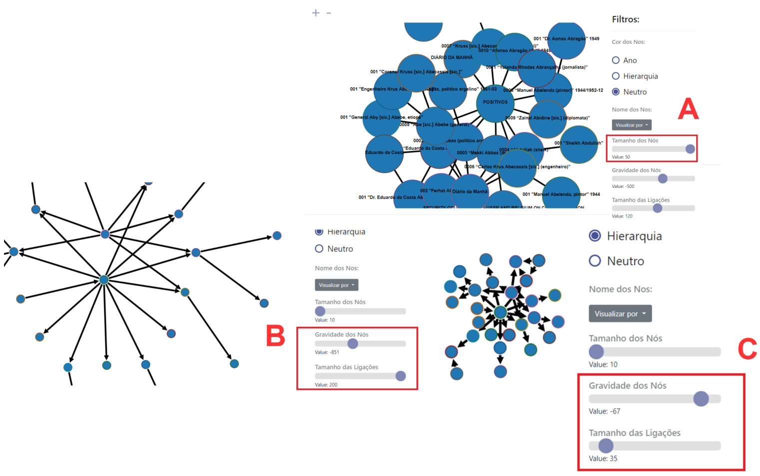


Figure 4.6: Change the dimensions of the graph.

regarding the amount of data available and displayed on the interface for better manipulation and comprehension.

4.2.3 Node information and visualization of linked data

If users want to see more information about a specific node, with 'double click' the application will open a modal with more information, Figure 4.7 shows all modals. The modal is divided into three labels, first label 'A' contains information about the record, second tab 'B' shows information about the other records that are linked to it. On tab 'C' users can see the organizations and actors that are connected and associated with the node. In the upper part of the modal, 'D', when clicking the icon book, users will have access to the page of the record as shown in Figure 4.8

4.2.4 Workflow and "how-to-use" features regarding the creation of nodes and links

4.2.4.1 Creation of new nodes:

After addressing the filters and main navigation features, the next functionality to mention is the creation of a new node. In Figure 4.9 is represented as the workflow of how to create a new node. In this example, 'A' when clicked in '+' a dropdown shows the type of node we want to create, 'B'. After selecting the node, it is automatically added to the graph representation 'C'.

4.2.5 Modify information node

After the creation of a new node, in Figure 4.10 users can change the undefined name to another one. For example, to 'Diogo Cunha'.

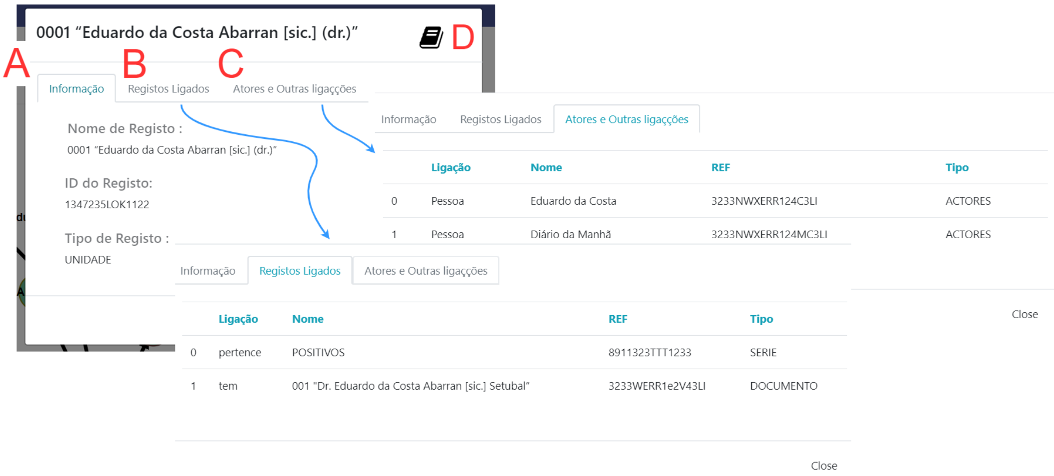


Figure 4.7: Modal of node with information about to the record.

Figure 4.8: Page of the record.

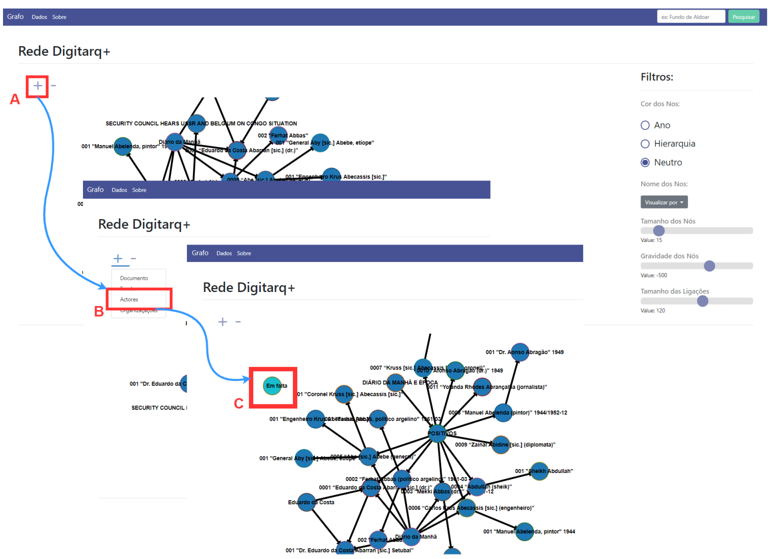


Figure 4.9: Creation of a new node into the graph.

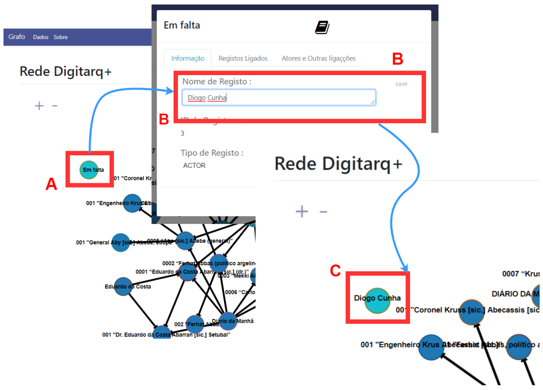


Figure 4.10: Edit the name of the node.

4.2.5.1 Create and delete links between historical records

For the navigation and to move nodes around the graph, users only need to click once in the node. But in order for the system to create a new link, users need to click the "CTRL" key and then click above the initial node and drag the arrow to the destination node. This will instantaneously create an arrow drawn from one to the other node. After the creation of the link, the system automatically adds the link to the table inside the modal, for example in Figure 4.11.

The same for deleting, users have to click the "CTRL" key for selecting the node or link, and then click the "delete" or "backspace" keys.

4.2.6 Conclusion

In this chapter, we presented a navigation graph tool for the visualization and manipulation of historical records. We have taken several aspects into consideration regarding the development of this web application. Aspects such as filtering tools to have a better comprehension of the visualization of the graph, the ability to interact with the nodes, move them around and see their information, are features that similar projects take into consideration, but not so many provide users the ability of manipulation.

One future problem, considering the migration from the ISAD(G) model to the CIDOC CRM model, is the fact that the volume of linked data will increase tremendously. Some solutions can

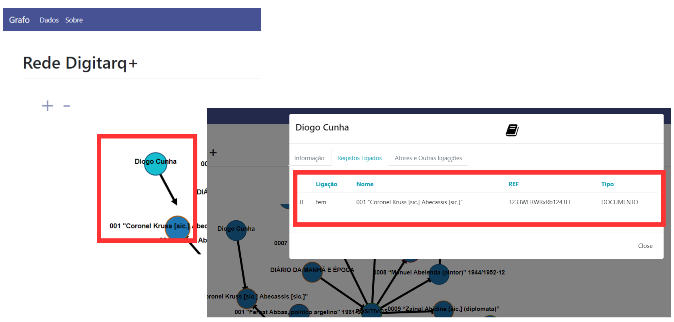


Figure 4.11: Creation of links between nodes.

consider what systems Captian and Heritamus used in their approach [17]. Instead of displaying all links, the system only shows the number of links entered by the user for each node.

Chapter 5

Conclusions and Future work

In this chapter we analyse the work done in this dissertation in terms of our goals, how we have accomplished them and the obtained results. We conclude with a discussion of future work that can be explored.

5.1 Balance of work done

During the period of development of this project, almost all the proposed objectives were fulfilled. The entire process of development was divided in two phases, as presented in Figure 5.1. The first phase focused on researching information and collecting solutions to the problem (Preliminary Work). The second phase focused on gathering requirements, development of a solution, and the evaluation of the implemented solutions, with the write of the final document.

In the Gantt diagram presented in Figure 5.1 is presented a detailed view of the work developed during this project. In the first phase, we analyzed the various topics to which the project was related to. At this stage, we have covered topics such as the Semantic Web; Linked-Data; how to visualize connected data; what ontologies are and how they are connected to the Semantic Web; graph representation techniques; and finally the study of some systems that were related to this

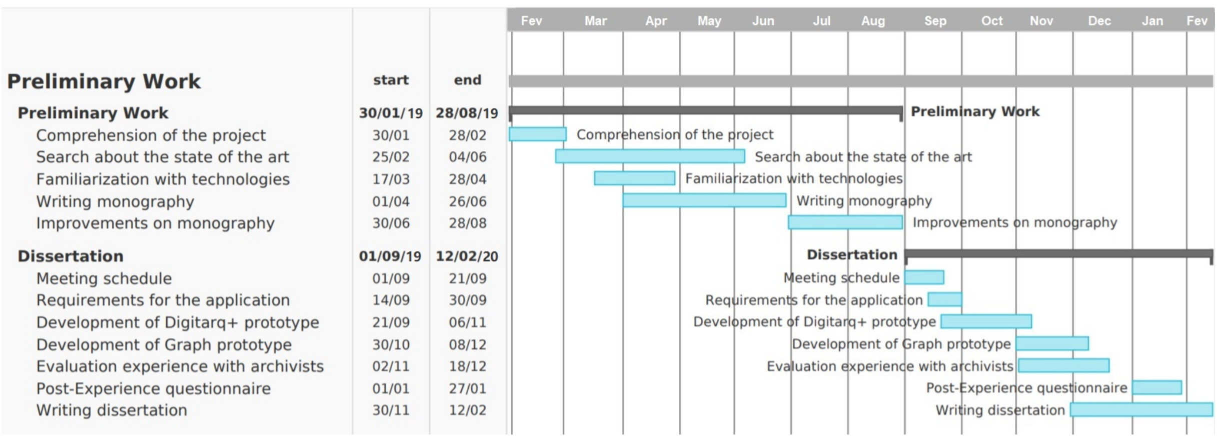


Figure 5.1: Gantt diagram.

areas to gather visual information available. In the second phase, we have analyzed technological solutions for the development of the web graph application and also the setup for the development of the prototype. In a third phase, the survey of requirements for the realization of the project was conducted, which involved the planning of the functionalities of the new DigitArq+ platform, the development of mock-ups for the interface and navigation flow between the existing pages. In the fourth phase, we have started the development of the DigitArq+ prototype, the implementation of the web graph application, and lastly, writing tests to evaluate the prototype. Finally, we have done the evaluation study with stakeholders with the implementation of the new DigitArq+ prototype, as well as the post-questionnaire.

5.2 Goals of the work

With our thesis, we aimed to contribute to the field of interaction with linked data, in particular in the context of historical archives. Firstly, we have analyzed the current state of the art about this topic and studied several research projects. We have analyzed semantic web applications and publications in the area. Then we analysed the old DigitArq systems to collected requirements for the new prototype. This study helped us understand what features could be changed, to provide different ways of interaction.

In our work, we have developed two applications to study. A new DigitArq+ non-functional prototype and a web graph application. The development of these platforms provides users the ability to navigate and explore real-world entities in historical archives, with new features to connect create linked data.

After the development, we have organized interviews with four DGLAB participants. To have a controlled environment during the interviews, we have used a script to guide the archivists while testing the prototype. With the recorded interviews using the concurrent think-aloud protocol, we have listed several topics which helped us to make conclusions and evaluate the collected data. Since we were using a non-functional prototype, we noticed some difficulties while retrieving the right data from the think aloud protocol. After these interviews, we conducted a questionnaire to ask participants their thoughts about the overall experience with the system.

This evaluation helped us to understand the participants' interests and difficulties in using the prototype. The study was divided into three stages: search, hierarchical, and creation of archival records. During the evaluation, we noticed that every time users were presented with new features, they tended to spend more time suggesting and comprehending the system. Moreover, some of the results were as expected, when users have performed slightly better the navigation with a graph perspective comparing to others, and the suggestions on the search feature were slightly more prominent as expected as well, due to the connected information presented on the interface.

With the evaluation, we also notice the interest of the participants to have the possibility of features to connect linked data and see easily other connected records. Besides the evaluation using the think aloud protocol, the result on the questionnaires was valuable to understand essential aspects. Besides the excellent feedback during the graph evaluation on the prototype and how friendly the navigation was, participants, on the other hand, during the questionnaire voted as it would be the less used feature for navigation and not so used during the creation. We think this is because users are familiar with the other features, making it easier for them to understand or to visualize how the navigation would perform.

5.3 Future work

The DigitArq+ prototype developed is an archival platform for DGLAB that was designed to support a linked data model to represent archival records. During the development we identified some features that would need more emphasis, such as the search filters where users mention the missed entities. The individual study and a deeper comprehension of these entities would also help us to see the differences between them. After these studies, the implementation and development of the prototype would be the next priority step to do, in order to provide the participants a more reliable and flexible system to navigate and interact with. After collecting these data, we now have a good funded base to develop a good system that goes exactly to the archivists' needs.

The navigation and manipulation of data within a graph perspective also caught participants' attention. Another future work idea of high priority that we couldn't do during this dissertation would be the study of our web graph application with real users. Participants tried the graph perspective for navigation and manipulation with the use of a non-functional prototype, and the next step would be for users to test with our web graph application — the evaluation study following with a post-experience questionnaire to gather more summarized information.

Appendix A

Digitarq+ Tasks Guide

Digitarq+

Guião Protótipo

Título: Guião de desenvolvimento para o protótipo não funcional Digitarq Desktop
Tempo estimado: 25-30min
Estudante: Diogo Almeida Cunha
Supervisor: Sergio Nunes
Data/Hora criação: 20 de Outubro, 2019
Formato: Word, PDF
Estatuto de utilização: Restrito ao grupo de trabalho ICON/EPISA Project

Diogo Almeida Cunha
up201405506@fe.up.pt

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Introdução

Neste guião é apresentado o layout e workflow do sistema **Digitarq+** através da apresentação de um conjunto de interfaces e interações implementadas num protótipo não-funcional. Com este guião pretende-se orientar a exploração deste protótipo tendo como objetivo principal recolher comentários dos utilizadores finais relativamente às opções seguidas no desenho das interfaces. O guião está estruturado em tópicos, que definem ações comuns. Para cada tópico é apresentada uma lista de passos que deve ser levada a cabo a partir da página (URL) indicada no início de cada tarefa. Durante a exploração de cada tarefa definida no guião, o utilizador deve explicitar em voz alta o raciocínio e as decisões tomadas em cada passo. Este método, baseado no protocolo "think-aloud", é usado para recolher dados que permitam informar o desenvolvimento futuro.

Índice

1. Login

Iniciar sessão no sistema Digitarq+
URL: /Login
1. Botão 'Sign Up' no Ceanto superior Direito
2. Botão 'GO'

(Login Iniciado)

2. Pesquisa

2.1 Pesquisa Hierárquica:

Pesquisa do doc. Simples '00001 Registo de Batismo 1811-07-07' pertencente ao Fundo Aldoar
URL: /PesquisaHierarquica
1. Clicar na seta do lado direito em 'Fundo da Paróquia Aldoar'
2. Clicar na primeira seta abaixo em Série '001 Registo de Batismo'
3. Clicar quarta seta abaixo, em 'Unidade 0004 Registo Batismo'
4. Abrir documento '00001 Registo de Batismo'
5. Navegar nas tabs: 'Descrição <-> Reserva <-> Permissões'

(Acesso a Documento por Hierarquia concluído)

2.2 Pesquisa através do Grafo

Pesquisa do doc. Simples '00001 Registo de Batismo 1811-07-07' pertencente ao Fundo Aldoar
URL: /PesquisaPorGrafo
1. Clicar 'pesquisa' canto superior esquerdo para retroceder
2. Clicar novamente em 'pesquisa' canto superior esquerdo
3. Fechar menu lateral, '3 barras' no canto superior esquerdo
4. Clicar 'all fields' em barra de pesquisa, topo da página
5. Clicar novamente em 'all fields' para seleccionar opção
6. Clicar funcionalidade 'Grafo'
7. Clicar em 'Search Here' para pesquisar por 'Fundo de Aldoar'
8. Clicar em 'Fundo de Aldoar' para iniciar pesquisa
9. Clicar em 'Loading' para iniciar navegação
   (navegação presente em CIDOC e modelo Hierárquico - filtros do lado direito)
10. Clicar nó 'Serie1'
11. Filtrar grafo, do lado direito clicar CIDOC CRM (ajuda a remover todas as ligações CIDOC).
12. Clicar nó 'Unidade 'Batismo 0004' (novamente para fechar se necessário)
13. Abrir documento '00001 Registo de Batismo 1811-07-07'
14. Navegar nas tabs: 'Descrição <-> Reserva <-> Permissões' (Acabando em Descrição)

(Acesso a Documento por Grafo concluído)

2.3 Pesquisa através de Texto

Pesquisa do doc. Simples '00001 Registo de Batismo 1811-07-07' pertencente ao Fundo Aldoar

URL: /PesquisaPorTexto

1. Clicar Em 'Grafo', navegação no topo da página
2. Clicar em 'Fundo de Aldoar' , navegação superior
3. Clicar em 'All Fields'
4. Clicar em 'Search Here' para pesquisar por 'Batismo de Aldoar'
5. Clicar em 'Batismo de Aldoar' para iniciar navegação

(Esquerdo: resultados dos arquivos; Meio: resultados relacionados; Direita: Filtragens)

6. Filtrar apenas por Documentos Simples, do lado direito
7. Abrir documento '0001 Registo de Batismo 1822-07-07'

(Acesso a Documento por Pesquisa de Texto concluído)

3. Registo Individual:

3.1 Adicionar Documentos Digitalizados:

Adicionar de forma rápida as últimas digitalizações feitas pelo utilizador ao documento;

URL: /AdicionarDigitalização

1. Clicar em 'Associar Últimas digitalizações?'

(Imagens adicionadas no canto superior direito)

3.2 Visualizar Imagem Arquivada

Funcionalidade de poder editar o documento e ver a imagem ao mesmo tempo;

URL: /RegistoIndividual/

1. Abrir imagem "Registo de Batismo" (terceira imagem do lado direito)
2. Clicar novamente em cima da imagem para expandir fotografia.
3. Clicar fora para sair de ecrã inteiro.

3.3 Editar Documento

Modificar o Título do registo de batismo para PT_CONTINENTE;

URL: /EditarDocumento

1. Clicar 'editar' em Identificação Arquivística
2. Editar Código do País, clicar no texto 'PT'
3. Guardar as alterações, clicar botão 'save'
4. Para minimizar imagem, clicar 'imagens_png' do lado direito em "Imagens Arquivadas"

3.4 Visualizar Pessoas Relacionadas

Ver pessoas que estão relacionadas com o Registo de Batismo

URL: /VerPessoasRelacionadas

1. Clicar na fotografia quadrada de 'Henrique VII' para ver outros documentos associados à mesma entidade (ps: clicar na fotografia, não no nome)
2. Clicar fora da caixa que abriu para sair

3. Clicar no nome ‘Henrique VII de Inglaterra’ para abrir informação da entidade
4. Clicar ‘editar’ caso seja pretendido alterar alguma informação
5. Clicar ‘save’ para guardar
6. Clicar em ‘leça do balio’ para ver nova entidade relacionada com Henrique IV
7. Retroceder, clicar seta canto superior esquerdo do modal
8. Clicar fora do *modal* para sair

3.5 Adicionar Pessoas Relacionadas

(9) Adicionar pessoas (já existentes) ao registo de Batismo
URL: /AdicionarPessoasRelacionadas
1. Clicar adicionar em Entidades Afiliadas (botão ‘+’ ao lado de *Maria I da Inglaterra*)
2. Procurar Entidade “Pedro Sebastião” , clicar barra search.
3. Selecionar ‘Pedro sebastião Andre’ no dropdown que apareceu.
4. Clicar na seta Afiliado, para escolher o tipo de entidade em relação ao documento
5. Selecionar ‘escritor’
6. Clicar ‘save’ para guardar entidade
(Nova entidade relacionada com o Registo de Batismo 00001 adicionada)

3.6 Criar Pessoas Relacionadas

Adicionar Pessoa ‘Afonso VI de Castela’ ainda não existentes no sistema para depois adicionar ao registo de batismo 00004
URL: /CriarPessoasRelacionadas
1. Clicar adicionar nova Entidade Afiliada (botão ‘+’) ao lado de Sebastião André
2. Clicar novamente no botão ‘+’ para abrir modal de criar entidade
3. Escrever nome ‘Afonso VI de Castela’, clicar em ‘adicionar nome’
4. Adicionar descrição da Entidade, clicar na caixa da descrição.
5. Adicionar ‘Leça do balio’ como entidade relacionada ao Afonso de Castela, clicar retângulo azul para completar a palavra.
6. Selecionar ‘Leça do Balio’
7. Clicar na entidade ‘Leça do balio’, para confirmar se a entidade está relacionada
8. Retroceder, seta no canto superior esquerdo
9. Clicar ‘search’ em Entidades Relacionadas para adicionar ‘Henrique de VI’ como irmão de Afonso VI de Castela
10. Selecionar ‘Henrique VI de Inglaterra
11. Associar Henrique VI como irmão de Afonso de Castela, clicar seta no bloco adicionado
12. Adicionar imagem, clicar na pasta do lado esquerdo do modal
13. clicar ‘save’ no canto superior direito
(Entidade criada e associada ao Registo de Batismoo 00001)
14. Clicar fora do modal para sair
(Podemos ver agora que a entidade ‘Afonso de Castela’ foi criada e adicionada ao registo)
15. Clicar ‘save’ e guardar a nova entidade adicionada ao Registo de batismo 0001
(Criação de uma nova Entidade conectada ao Registo de Batismo 0004 concluída)

3.7 Editar Descrição

Agora vamos editar a descrição do mesmo registo, para adicionar a entidade 'Afonso de Castela' que acabamos de criar

URL: /EditarDescrição

1. Ver tab 'Contexto Arquivista', clicar na seta do lado direito
2. Clicar 'editar'
3. Clicar '....' para adicionar novo texto descritivo com ''Afonso de Castela' associado

(retângulo azul significa que a entidade foi reconhecida pelo sistema)

4. Clicar em 'Afonso de Castela' para selecionar a entidade pretendida
5. Clicar 'save' depois da entidade estar associada.
6. Clicar na entidade 'Afonso de VI' para navegar até à entidade
7. Clicar fora do modal para sair

3.8 Funcionalidade Conservar

Forma mais rápida de aceder à conservação. Visualizar do *Registo Batismo 0004 de 1812-02-02* as recentes conservações efetuadas pelo utilizador

URL: /ConservarDocumento

1. Clicar em 'Pesquisa' para retroceder, canto superior esquerdo
2. Clicar em 'All flieds'
3. Clicar novamente em 'all field' para listar as funcionalidades
4. Selecionar ação de 'Conservar'
5. Adicionar na barra o código de referência do arquivo, clicar em 'Search Here'
6. Clicar no código de ref. para avançar
7. Clicar em 'Conserva_ARX_2' para ver mais informação sobre a alteração
8. Podemos ler em 'descrição modificada' o que foi corrigido ou alterado na conservação.
9. Clicar novamente em 'Conservar_ARX_2' para encolher opção.

(Visualização das considerações efetuadas no Registo Batismo 0004 concluído)

4. Criação :

4.1 Individual/Hierárquica

Adicionar um documento simples à série 001 Registo de batismo 1644-02, com auto-complete de informação associada aos documentos anteriores

URL: /CriarDocumento

1. Clicar na tab 'Descrição'
2. Clicar em 'Pesquisa', canto superior esquerdo
3. Diminuir Hierárquica, clicar na seta para cima em 'Unidade 0004 Registo Batismo 2844-0707'
4. Clicar adicionar , botão '+' do lado esquerda da seta
5. Selecionar opção 'Documento'

6. Para apagar documento, clicar ‘x’ em cima de ‘digitalizar’
7. Criar novamente um documento novo

4.1.1 Autocomplete

(14) Tornar mais rápido e eficaz a criação de documentos ligados
URL: /FuncionalidadeAutoComplete
1. Utilizar a funcionalidade *‘auto-complete’* que o sistema fornece. Lado direito
2. Em caso da informação estar errada, clicar ‘desfazer auto-complete’
3. No seguinte caso, usar a informação do auto-complete e clicar novamente na funcionalidade
4. clicar em ‘editar’, ao lado de Identificação Arquivística
5. Clicar no nome do registo para modificar o Nome
6. Adicionar ‘Contexto Arquivista’ , clicar na seta lado direito para abrir painel
7. Clicar ‘adicionar descrição’ para escrever o texto
8. Guardar as alterações efetuadas, clicar save do lado direito de ‘Contexto Arquivista’
9. Clicar na navegação em ‘unidade 0004’ no topo da página , para ver ficheiro adicionado

(Documento adicionado à Unidade de instalação 0004 concluído)

4.2 Criação em Grafo

Adicionar um documento simples no fundo de Aldoar, série 06, através do grafo
URL: /CriaçãoPorGrafo
1. Clicar ‘pesquisa’ no canto superior esquerdo
2. Minimizar dashboard, clicar 3 barras no canto superior esquerdo
3. Selecionar ‘all fields’
4. Clicar em ‘ all fields’ e escolher opção ‘Grafo’
5. Clicar em ‘search here’ para introduzir o nome do documento
6. Clicar em ‘fundo de aldoar’
7. Clicar em ‘loading’ para iniciar o grafo
8. Expandir nó ‘Serie1’
9. Clicar em cima do nó ‘Serie6’
10. Clicar na opção ‘Nó’ para visualizar modal com informação sobre o nó

(Podemos visualizar na lista de propriedades, que é composto por 2 ligações)

11. Clicar fora do modal para sair

4.2.1 Adicionar nó através do Pai Pretendente

Criar um nó ‘Batismo de Aldoar 0002’, inicialmente já conectado ao nó pai;
URL: /AdicionarNoAtravesPai
1. Clicar novamente no nó ‘Serie6’
2. Selecionar opção ‘Criar’
3. Selecionar o tipo de ligação, ‘Pertence’

(Nome automaticamente adicionado)

4. Abrir nó ‘Batismo de aldoar 0002’
(podemos ver nas propriedades que já tem uma ligação atribuída ao próprio nó)
5. Aceder ao documento, clicar no livro canto superior direito do modal
6. Clicar em ‘Grafo’ para retroceder no canto superior esquerdo
7. Clicar em ‘save changes’

4.2.2 Remover Nós e Links

Remover a ligação do nó ‘Batismo de Aldoar 0002’ acabado de criar;
URL: /RemoverLigação
8. Clicar tesoura no canto superior esquerdo
9. Clicar em cima da ligação entre ‘Serie06 <-> batismo aldoar 0002’ para quebrar relação

4.2.3 Adicionar nó individual (permite criar Fundos)

Criar um nó ‘Batismo de Aldoar 0002’, inicialmente sem estar ligado ao nó pai;
URL: /AdicionarNoSemPai
1. Clicar no botão adicionar (botão ‘+’ canto superior esquerdo)
2. Selecionar ‘Unidade’ Simples para o tipo de documento
(nome nao adicionado desta vez por não estar ligado ao nó pai)
3. Abrir o novo nó adicionado
4. Inserir nome do nó, clicar em ‘Inserir nome…’
5. Clicar ‘Add property’ , para adicionar nova ligação ao nó
6. Clicar em ‘Serie06’ , para criar ligação entre os nós pai e filho
7. Escolher o tipo de ligação, selecionar ‘pertence’
8. Selecionar nó ‘batismo de aldoar’ e visualizar a nova ligação concluída com o valor ‘pertence’ em relação à ‘serie6’
9. Clicar ‘save’ para guardar modificações
(Ação de adicionar documentos e criar ligações concluída)

4.2.4 Criar e adicionar novas Entidades

Adicionar a entidade ‘Afonso de Vi de Inglaterra’ a outros três nós já existentes;
URL: /CriarEntidadePorGrafo
1. Clicar botão adicionar , canto superior esquerdo botão ‘+’
2. Selecionar ‘Entidade’
3. Abrir nó da Entidade, clicar no nó amarela lado esquerdo
4. Escrever o nome da Entidade ‘Afonso de VI de Inglaterra’ , clicar em ‘Inserir nome’
5. Clicar em ‘add property’ para adicionar documentos associados à mesma Entidade.
6. Selecionar nó ‘batismo de aldoar’ (no roxo canto superior esquerdo)
7. Selecionar nó ‘serie2’ (no azul lado direito)
8. Selecionar nó ‘Formal’ (no azul lado esquerdo)
9. Clicar ‘Parar de Ligar Links’ para parar de adicionar nós à entidade
10. Podemos visualizar na lista, os nós que foram acabados de serem conectados
11. Clicar icon User no canto superior direito do modal para ver informação sobre a entidade
12. Clicar na seta para retroceder, canto superior esquerdo do modal

13. Clicar fora do modal para sair.
14. Clicar no utilizador autenticado ‘José Artur’, canto superior direito, para sair do sistema
(Ligação entre entidades e documentos através do grafo concluída)

Guião Concluído

Appendix B

Questionnaire about UX Sessions

Digitarq+ :: Questionário Sessões UX

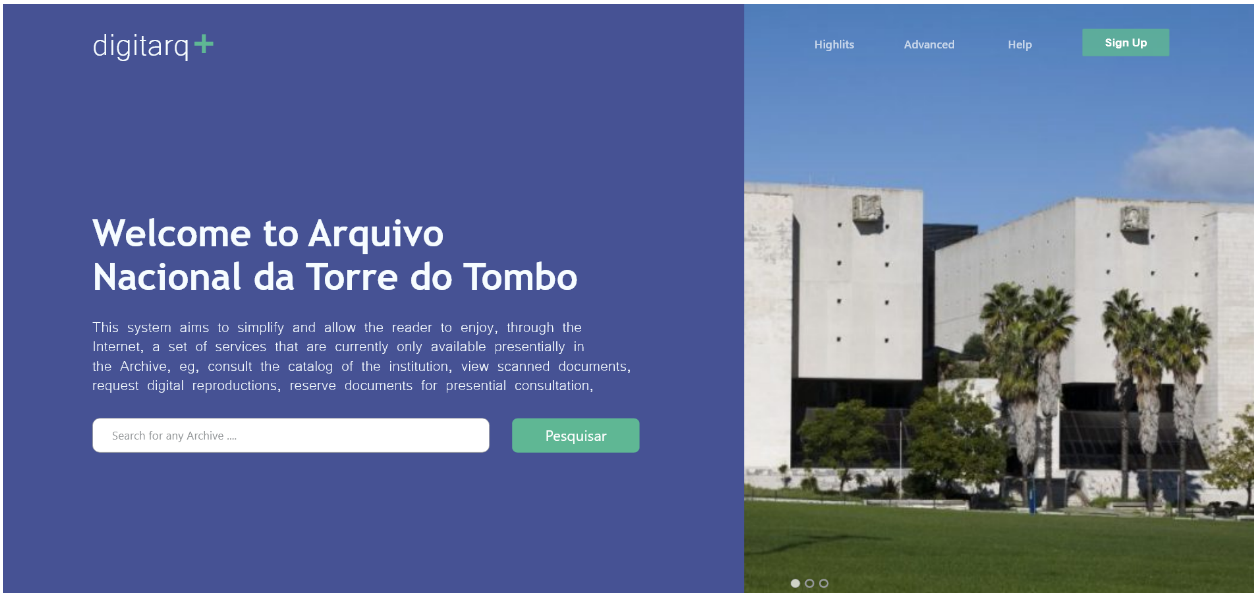
Este questionário tem como objetivo recolher informações após as sessões realizadas sobre o protótipo não-funcional do Digitarq+. As imagens servem de enquadramento às perguntas em questão. Caso seja necessário, poderá aceder ao guião e ao protópito através das seguintes ligações.

Guião: https://drive.google.com/file/d/10XE0M7hZadkwcOkRMd0nnaA4nHoYIY0I/view

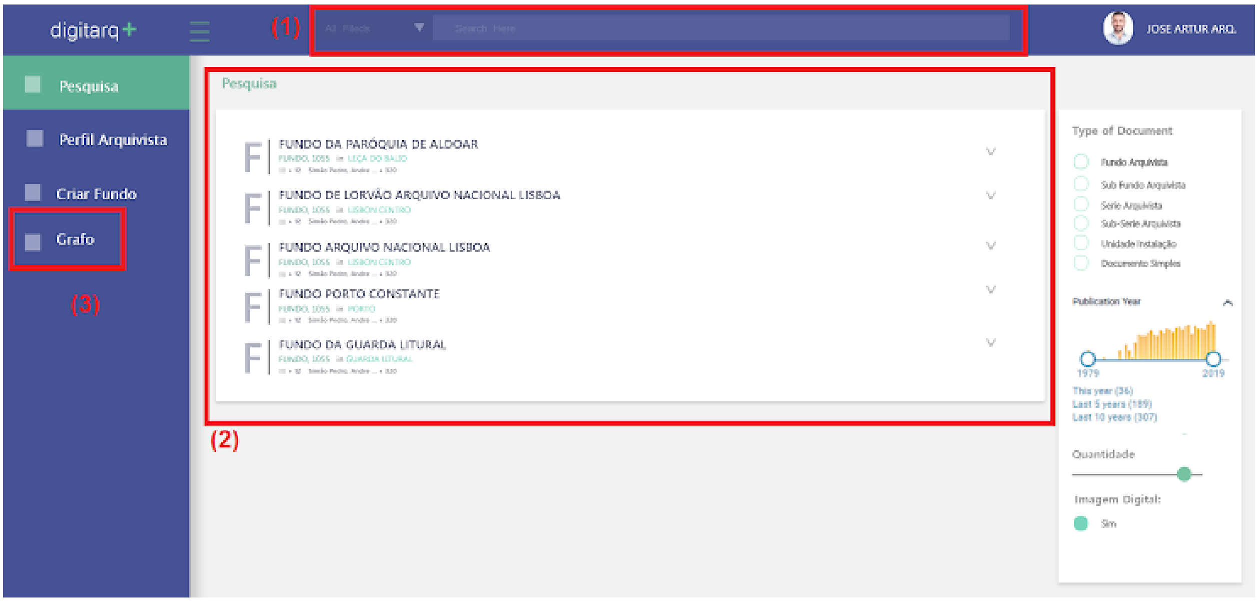
Protótipo: https://xd.adobe.com/spec/61090899-20d6-4b5f-4465-946ec41caa09-1738/?fullscreen

*Obrigatório

Digitarq+ Página Inicial



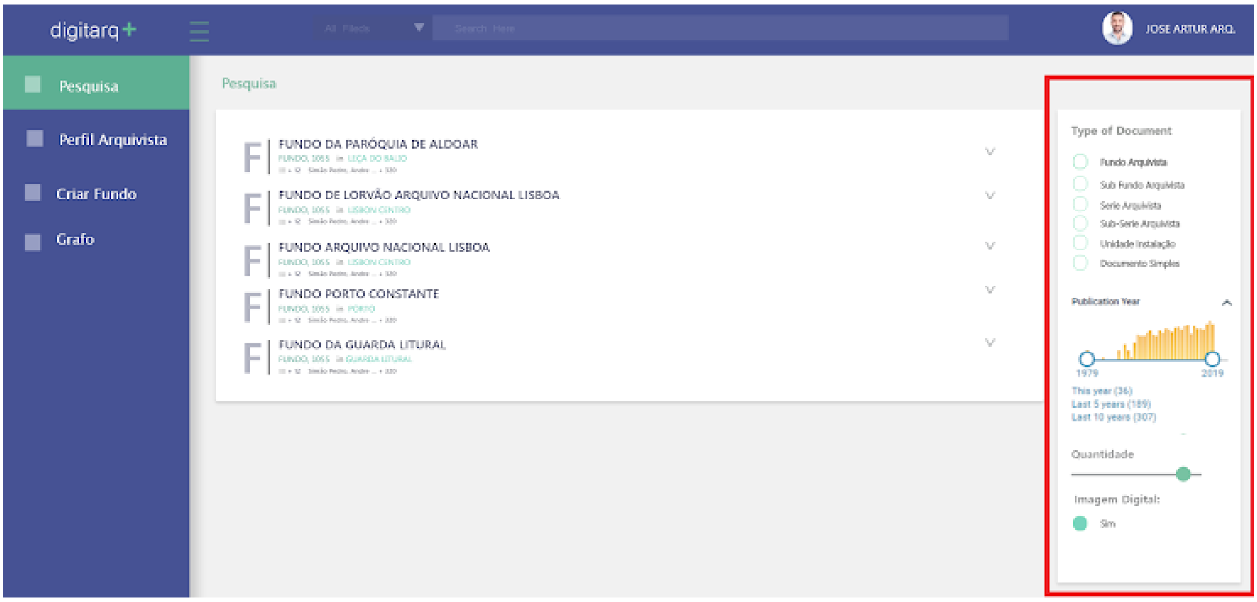
1. 1. [Pesquisa] Sabendo que o sistema Digitarq+ pode ser utilizado para pesquisa e navegação de registos, indique qual o método que considera ter maior potencial de uso: *



Marcar apenas uma oval por linha.

	Sem Utilização	Utilização rara	Utilização nomal	Utilização frequente
Acesso atráves de pesquisa textual (1)				
Acesso através da hierarquia (2)				
Acesso através do grafo (3)				

2. 2. [Filtragem] Para a filtragem dos registos encontrados, o sistema poderá fornecer diferentes mecanismos. Indique o grau de interesse para cada um dos seguintes tipos de filtro: *



Marcar apenas uma oval por linha.

	Sem utilização	Pouco relevante	Relevante	Muito relevante
Tipo de documento (ex: fundo, série, etc)				
Ano de publicação				
Quantidade de documentos				
Documentos relacionados				
Documentos com imagem				
Associados a localidade geográfica				
Associados a tipo de instituição				
Associados a pessoas				

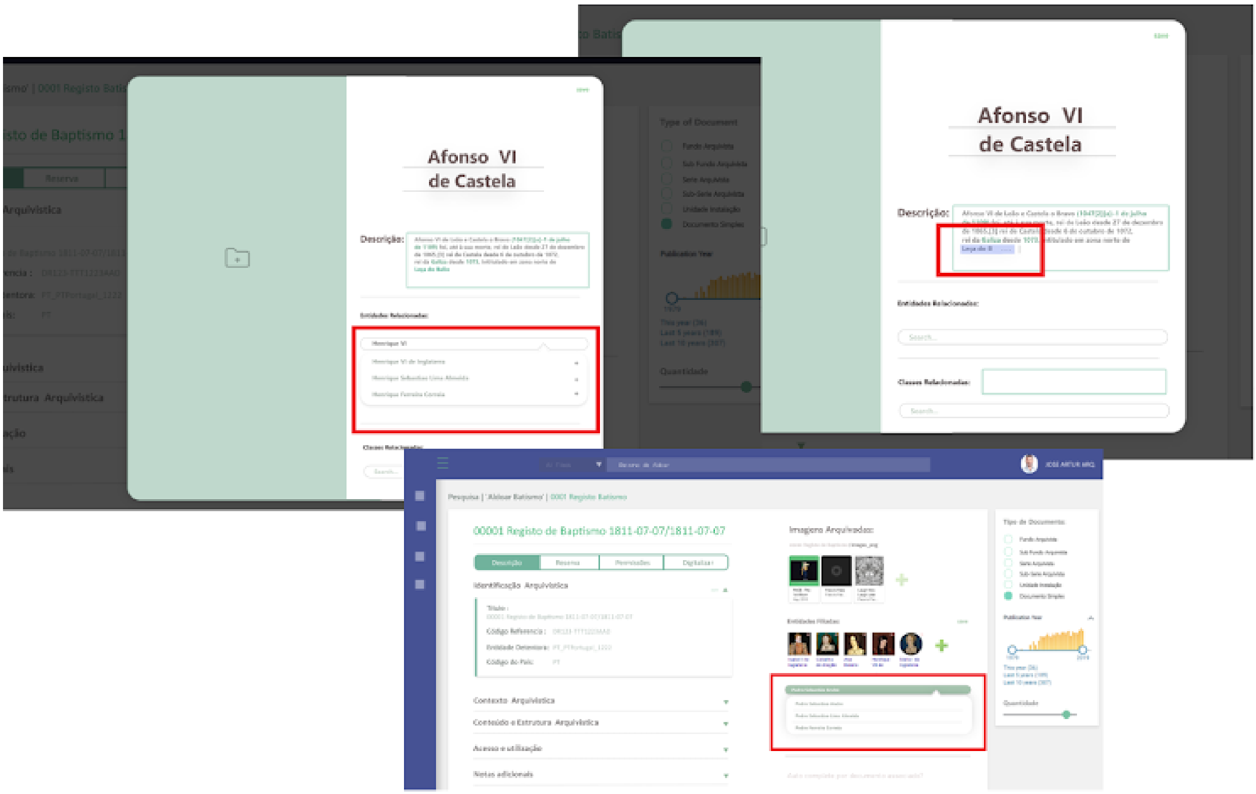
3. 3. [Filtragem] Acha os critérios apresentados antes suficientes? Se 'sim', pode avançar. Se 'não', pode apresentar 'outras opções'. *

Marcar tudo o que for aplicável.

- [ ] Sim, suficientes.

Outra: [ ] ______________________

4. 4. [Ligações] Na construção de ligações entre documentos e entidades (pessoas, locais, etc), qual será a solução preferencial para identificar as entidades alvo? *



Marcar apenas uma oval por linha.

	Discordo totalmente	Discordo parcialmente	Concordo parcialmente	Concordo totalmente
Através de uma janela de navegação	⬭	⬭	⬭	⬭
Usando um código de referência	⬭	⬭	⬭	⬭
Através de uma lista	⬭	⬭	⬭	⬭
Através de navegação em árvore	⬭	⬭	⬭	⬭

5. 5. Para a mesma questão, deixamos uma resposta em aberto para sugestões ou comentários sobre a contrução de ligações com entidades:

6. 6. O novo sistema Digitarq+, ao seguir o modelo CIDOC CRM, irá permitir a ligação de cada registo com diferentes entidades. Ordene, por ordem de importância, as entidades que considera mais relevantes neste contexto: *

Marcar apenas uma oval por linha.

	Sem relevância	Pouco relevante	Indiferente	Relevante	Mais relevante
Pessoas					
Organizações					
Datas					
Locais					

7. 7. Para a mesma questão, deixamos uma resposta em aberto para mais sugestões relacionadas com entidades: *

8. 8. Para as seguintes tarefas, qual a importância de ter o garfo explicitamente visível nas seguintes operações:

Marcar apenas uma oval por linha.

	Discordo totalmente	Discordo parcialmente	Concordo parcialmente	Condordo totalmente
Pesquisa e navegação				
Criação e edição de registos				
Consulta de registos				

9. 9. De um modo geral, após a utilização do prototipo Digitarq+, qual a sua opinião geral relativamente à navegação e interface com o utilizador do sistema: *

Marcar apenas uma oval.

- Muito abaixo das expectativas
- Abaixo das expectativas
- Acima das expectativas
- Muito acima das expectativas

10. 10. Considerando uma perspetiva geral sobre o protótipo experimentado, indique outros aspetos ou comentários que considera importante registar: *

References

[1] Martim de Albuquerque. A torre do tombo e os seus tesouros. *Inapa Journal- Lisbon*, 1990.

[2] Anton Anikin, Dmitry Litovkin, Marina Siplivaya, Elena Sarkisova, and Tatyana Petrova. Ontology visualization: Approaches and software tools for visual representation of large ontologies in learning. In *Ontology Visualization: Approaches and Software Tools for Visual Representation of Large Ontologies in Learning*, pages 133–149. Interactive Visualization of Ontology-Based Conceptual Domain Models in Learning and Scientific Research, 08 2017.

[3] Dr AnitaKumari and Jawahar Thakur. Semantic web search engines: A comparative survey. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 2019.

[4] Tim Berners-Lee. Semantic web on xml. *Keynote presentation for XML*, 2000, 2000.

[5] Tim Berners-Lee, Dimitri Dimitroyannis, A John Mallinckrodt, and Susan McKay. World wide web. *Computers in Physics*, 8(3):298–299, 1994.

[6] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts*, pages 205–227. tim2, 2011.

[7] Tim Bray, Jean Paoli, C Michael Sperberg-McQueen, Eve Maler, and Franois Yergeau. Extensible markup language (xml) 1.0. *W3C Cecommendation October*, 2000.

[8] Balakrishnan Chandrasekaran, John R Josephson, and V Richard Benjamins. What are ontologies, and why do we need them? *IEEE Intelligent systems*, 1:20–26, 1999.

[9] Rohrer Christian. When to use which user-experience research methods. `https://www.nngroup.com/articles/which-ux-research-methods/`. Accessed: 2014-09-30, 2014.

[10] CRM CIDOC. The cidoc conceptual reference model, 2003.

[11] Mariano P Consens. *Creating and filtering structural data visualizations using Hygraph patterns*. University of Toronto, 1994.

[12] Joseph S Dumas, Joseph S Dumas, and Janice Redish. *A practical guide to usability testing*. Intellect books, 1999.

[13] K Anders Ericsson and Herbert A Simon. How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, 5(3):178–186, 1998.

[14] Fredo Erxleben, Michael Günther, Markus Krötzsch, Julian Mendez, and Denny Vrandečić. Introducing wikidata to the linked data web. In *International Semantic Web Conference*, pages 50–65. Springer, 2014.

[15] Pedro Félix. Heritamus: A tool for collaborative curation of tangible and intangible heritage. *Transposition. Musique et Sciences Sociales*, page 8, 2019.

[16] Miguel Ferreira and José Carlos Ramalho. Digitarq: creating a historical digital archive. thesis, UMINHO, 2004.

[17] Fatma Ghorbel, Nebrasse Ellouze, Elisabeth Métais, Fayçal Hamdi, Faiez Gargouri, and Noura Herradi. Memo graph: an ontology visualization tool for everyone. *Procedia Computer Science*, 96:265–274, 2016.

[18] Leticia Ruiz Gómez. The prado musuem: A spanish perspective. In *The Majesty of Spain: Royal collections from the Museo del Prado and the Patrimonio Nacional*, pages 20–22. The Mississippi Commission for International Cultural Exchange, 2001.

[19] Ivan Herman, Guy Melançon, and M Scott Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on visualization and computer graphics*, 6(1):24–43, 2000.

[20] B. Chandrasekaran John R. Josephson and V. Richard Benjamins. What are ontologies, and why do we need them. *IEEE Intelligent systems*, pages 20–26, 01 1999.

[21] Thomas Kamps, Joerg Kleinz, and John Read. Constraint-based spring-model algorithm for graph layout. In *International Symposium on Graph Drawing*, pages 349–360. Springer, 1995.

[22] G Kuck. Tim berners-lee's semantic web. *SA Journal of Information Management*, 6(1), 2004.

[23] Catherine C Marshall and Frank M Shipman. Which semantic web? In *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, pages 57–66. ACM, 2003.

[24] Budiu Raluca. Quantitative vs. qualitative usability testing. `https://www.nngroup.com/articles/quant-vs-qual/`. Accessed: 2014-09-30, 2014.

[25] Elizabeth Shepherd and Charlotte Smith. The application of isad(g) to the description of archival datasets. *Journal of the Society of Archivists*, 21(1):55–86, 2000.

[26] Arnold POS Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries*, pages 521–530. ACM, 2010.

[27] W3. Web semantic page. `https://www.w3.org/2001/sw/wiki/Marbles`. Accessed: 2009-05-29, 2009.